

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: March 6, 2006, 19:48:01 ; Search time 41.6 Seconds
(without alignments)
2323.495 Million cell updates/sec

Title: US-09-360-685C-26
Perfect score: 1171
Sequence: 1 aaaaatgcaataaaga.....caatcaagcgatcagaag 685

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_plus.n2p.model -DEV=rlp
-Q=/abs/ABSSWEB.spool/US09360685/runat_06032006_121316_15362/app_query.fasta_1
-DB=UniProt -QFMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOFC=0 -LOOEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p
-USER=US09360685@CCN_1_1_466@runat_06032006_121316_15362 -NCPU=6 -ICPU=3
-NO_MMAP -NBS_SCORES=0 -WAIT -DSRLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: UniProt_05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1127	96.2	1147	1	CGA1_HELPY
2	1127	96.2	1147	1	CGA1_HELPY
3	1035	88.4	1198	2	Q6VRO5_HELPY
4	1029	87.9	1181	2	Q6VRW7_HELPY
5	1029	87.9	1230	2	Q4PLI5_HELPY
6	1025.5	87.6	1230	2	Q4PLI5_HELPY
7	1016	86.8	1142	2	Q8RNU0_HELPY
8	1009	86.2	1180	2	Q52GZ6_HELPY
9	1004	85.7	1179	2	Q8RZK3_HELPY
10	1003.5	85.7	1179	2	Q8RZK3_HELPY
11	992	84.7	1186	2	Q52GZ6_HELPY
12	990.5	84.6	1179	2	Q52GZ6_HELPY
13	984	84.0	1182	2	Q6VRG7_HELPY
14	982.5	83.9	1394	2	Q8RRY0_HELPY
15	972.5	83.0	327	2	Q8RZB0_HELPY
16	972.5	83.0	1148	2	Q60FP6_HELPY

17	963	82.2	441	2	Q917K3_HELPY	Q917K3 helicobacte
18	962.5	82.2	359	2	Q8RZB8_HELPY	Q8RZB8 helicobacte
19	962.5	82.2	1183	2	Q60FQ5_HELPY	Q60FQ5 helicobacte
20	960	82.0	1183	2	Q60FQ5_HELPY	Q60FQ5 helicobacte
21	959.5	81.9	1186	1	CAGA_HELPY	CAGA_HELPY
22	958.5	81.9	1247	2	Q9F223_HELPY	Q9F223 helicobacte
23	958.5	81.9	1247	2	Q9F223_HELPY	Q9F223 helicobacte
24	955.5	81.6	1156	2	Q5D6R2_HELPY	Q5D6R2 helicobacte
25	952	81.3	1216	2	Q8RZK9_HELPY	Q8RZK9 helicobacte
26	948.5	81.0	1247	2	Q07910_HELPY	Q07910 helicobacte
27	944.5	80.7	408	2	Q917K4_HELPY	Q917K4 helicobacte
28	942.5	80.5	1184	2	Q60FQ2_HELPY	Q60FQ2 helicobacte
29	941	80.4	1222	2	Q8RZB8_HELPY	Q8RZB8 helicobacte
30	939.5	80.2	359	2	Q8RZB8_HELPY	Q8RZB8 helicobacte
31	939.5	80.2	1183	2	Q60FP5_HELPY	Q60FP5 helicobacte
32	939.5	80.2	1183	2	Q60FP7_HELPY	Q60FP7 helicobacte
33	939.5	80.2	1183	2	Q60FP3_HELPY	Q60FP3 helicobacte
34	939.5	80.2	1188	2	Q60FN9_HELPY	Q60FN9 helicobacte
35	938.5	80.1	1190	2	Q5D6R1_HELPY	Q5D6R1 helicobacte
36	937.5	80.1	1183	2	Q60FP9_HELPY	Q60FP9 helicobacte
37	934	79.8	391	2	Q8RZC1_HELPY	Q8RZC1 helicobacte
38	934	79.8	1214	2	Q60FQ8_HELPY	Q60FQ8 helicobacte
39	930.5	79.5	1183	2	Q8RZB5_HELPY	Q8RZB5 helicobacte
40	930.5	79.5	1183	2	Q60FQ4_HELPY	Q60FQ4 helicobacte
41	930.5	79.5	1189	2	Q60FP4_HELPY	Q60FP4 helicobacte
42	926.5	79.1	1320	2	Q6VRU6_HELPY	Q6VRU6 helicobacte
43	918	78.4	1184	2	Q60FP2_HELPY	Q60FP2 helicobacte
44	912.5	77.9	445	2	Q917K5_HELPY	Q917K5 helicobacte
45	902.5	77.1	1182	1	CGA2_HELPY	CGA2_HELPY

ALIGNMENTS

RESULT 1

CGA1_HELPY STANDARD; PRT: 1147 AA.
ID CGA1_HELPY
AC P80200.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Cytotoxicity associated immunodominant antigen (120 kDa protein).
GN Name=cga; Synonyms=cag;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CTUG 17874 / NCTC 11638;
RX MEDLINE=93296225; PubMed=8516329;
RA Covacci A., Censini S., Bugnoli M., Petraccia R., Burroni D.,
RA Macchia G., Massoni A., Papini E., Xiang Z., Figura N., Rappunli R.;
RT "Molecular characterization of the 128-kDa immunodominant antigen of
RT Helicobacter pylori associated with cytotoxicity and duodenal ulcer";
RL Proc. Natl. Acad. Sci. U.S.A. 90:5791-5795(1993).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CTUG 17874 / NCTC 11638;
RX MEDLINE=97121442; PubMed=8962108; DOI=10.1073/pnas.93.25.14648;
RA Censini S., Lange C., Xiang Z., Cradocke J., Ghara P., Borodovsky M.,
RA Rappunli R., Covacci A.;
RT "cag, a pathogenicity island of Helicobacter pylori, encodes type I -
RT specific and disease-associated virulence factors";
RL Proc. Natl. Acad. Sci. U.S.A. 93:14648-14653(1996).
RN [3]
RP PROTEIN SEQUENCE OF 181-190; 314-328; 366-377; 420-430; 494-506;
RP 661-677; 900-914 AND 1062-1077.
RA Herrmann V., Herrmann J., Kist M.;
RL Submitted (APR-1993) to Swiss-Prot.
CC -!- FUNCTION: May be necessary for the transcription, folding, export,
CC or function of the cytotoxin.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.

DR EMBL; X70039; CAA49633.1; -; Genomic DNA.
DR EMBL; AF282853; AAC44706.1; -; Genomic DNA.
DR PIR; B48281; B48281.
DR InterPro; IPR005169; Caga.
DR InterPro; IPR004355; IVSec_caga.
DR Pfam; PF03507; Caga; 1.
DR PRINTS; PR01553; TYPE4SSCAGA.
KM Antigen; Direct protein sequencing.
FT COMPBIAS 880 885 Poly-Aan.
FT CONFLICT 320 320 G -> A (in Ref. 3).
FT CONFLICT 325 325 P -> F (in Ref. 3).
FT CONFLICT 328 328 R -> K (in Ref. 3).
FT CONFLICT 426 426 K -> E (in Ref. 3).
FT CONFLICT 429 429 O -> E (in Ref. 3).
FT CONFLICT 673 675 AQN -> TMD (in Ref. 3).
FT CONFLICT 901 901 A -> T (in Ref. 3).
FT CONFLICT 903 903 Q -> E (in Ref. 3).
FT CONFLICT 907 907 L -> R (in Ref. 3).
FT CONFLICT 910 910 P -> R (in Ref. 3).
FT CONFLICT 914 914 O -> R (in Ref. 3).
FT CONFLICT 1072 1072 P -> S (in Ref. 3).
FT CONFLICT 1074 1074 S -> D (in Ref. 3).
SQ SEQUENCE 1147 AA; 128014 MW; AB92770835f68490 CRC64;

Alignment Scores:
Pred. No.: 1 61e-72
Score: 1127.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 96.2%
DB: 1
Gaps: 0

US-09-360-685c-26 (1-685) x CGAL_HELPY (1-1147)

QY 1 AAAAAAGCAAAATTAAGATTTTCAGCAAGGTAAACCAAGCAAAAGCACTTGAAT 60
Db LybhangilyshamlyshapheSerlyvalthrglnalysSerapleuGlu 769
QY 61 TCCGTTAAAGATGATCATCAATCAAAAGGTAAAGTTGATATCTCAATCA 120
Db SerVallyshapValilleleanglnlyvalthraspLyvalaephenuleuGln 789
QY 121 GCGGTATCAGTGCTTAAAGCAAGGTGATTTTCAGTGGTGAAGCAAGCTTACCGAT 180
Db AlavalSerValalalyvalathrglyaspPheSerAlyvalGluGlnalaleuhalasp 809
QY 181 CTCAAAAATTTCTCAAAAGCAATTGCCCAACAGCTCAAAAATTAAGATCTCAAT 240
Db LeuLyshapheSerlysglnleuhalaglnGlnalaglnlyshamlyshSerleuasn 829
QY 241 GCTAGAAAAAATGTAATATATCAATCCGTTAAGATGGTGAATGGAACCTTGAAT 300
Db AlathrglyshySerGlnlylethrglnSerValysanglyvalandlythleuVal 849
QY 301 GGTAAAGGTTATCTCAAGCAGAGCAACCTTTCTTAAAACTTTCCGACATCAAG 360
Db GilyaenglyleuSerGlnlaglualathrthleuSerlyshamlyshSerleuasn 869
QY 361 AAGAGTTGAATGCAAACTTGAAATTTCAATTAACAATTAAGCACTTGAATCAAAAC 420
Db LysGlnleuasnAlalyseuGlyshPheasnashamlyshamlyshleuLyshasn 889
QY 421 GAACCAATTTATGTAAGTAATTAAGAAAGAGAGGAGAGCAAGCTTACCTTGAAGA 480
Db GlnPhelethryalalyvalaenlyshyvalaglyglnhalasSerleuGln 909
QY 481 CCCATTACGCTCAAGTGTCAAAAAGGTAAATGCAAAAATGACCCGATCAATCAATA 540

Db 910 ProthryalaglnValalalyshyvalaenhalalythleuaspAlyleuGlnle 929
QY 541 GCAAGTGTGGGTTGAGGCAAGCAAGCGGCTTCCCTTGAAGAAAGCATTAATA 600
Db AlaserlyleuGlnlyvalalalyglnhalaleuPheleuLyshyghlyshaplys 949
QY 601 GTGATGATCTCAAGTGAAGGCTTTCAGAGATCAAGAAATGGCTCAGAAAATTGAC 660
Db ValasphleuSerlyshyvalGlyleuSerlyshanglnleuhalaglnlyshleuasp 969
QY 661 AATCTCAATCAAGCGCTTTCAGAA 684
Db 970 AenleuasnGlnhalalaserGln 977

RESULT 2
Q9P220_HELPY
ID Q9P220_HELPY PRELIMINARY; PRT; 1147 AA.
AC Q9P220_
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Caga.
GN Name=caga,
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxId=210;

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCCTC1637;
RX MEDLINE=21457648; PubMed=1573724; DOI=10.1007/s00530070002;
RA Hosono F.B., Katsuyama K., Watanabe K., Takahashi S., Uchimura H.,
RT Ando T.;
RT "Heterogeneity found in the caga gene of Helicobacter pylori from
RT Japanese and non-Japanese isolates.";
DR Gaetochentrol. 35:890-897(2000).
DR EMBL; AB015416; BAB20926.1; -; Genomic DNA.
DR GO; GO:0019534; F:toxin transporter activity; IEA.
DR InterPro; IPR005169; Caga.
DR InterPro; IPR004355; IVSec_caga.
DR Pfam; PF03507; Caga; 1.
DR PRINTS; PR01553; TYPE4SSCAGA.
SQ SEQUENCE 1147 AA; 128003 MW; D740K7275A9B141D CRC64;

Alignment Scores:
Pred. No.: 1 61e-72
Score: 1127.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 96.2%
DB: 2
Gaps: 0

US-09-360-685c-26 (1-685) x Q9P220_HELPY (1-1147)

QY 1 AAAAAAGCAAAATTAAGATTTTCAGCAAGGTAAACCAAGCAAAAGCACTTGAAT 60
Db LybhangilyshamlyshapheSerlyvalthrglnalalySerapleuGlu 769
QY 61 TCCGTTAAAGATGATCATCAATCAAAAGGTAAAGTTGATATCTCAATCA 120
Db SerVallyshapValilleleanglnlyvalthraspLyvalaephenuleuGln 789
QY 121 GCGGTATCAGTGCTTAAAGCAAGGTGATTTTCAGTGGTGAAGCAAGCTTACCGAT 180
Db AlavalSerValalalyvalathrglyaspPheSerAlyvalGluGlnalaleuhalasp 809
QY 181 CTCAAAAATTTCTCAAAAGCAATTGCCCAACAGCTCAAAAATTAAGATCTCAAT 240
Db LeuLyshapheSerlysglnleuhalaglnGlnalaglnlyshamlyshSerleuasn 829
QY 241 GCTAGAAAAAATGTAATATATCAATCCGTTAAGATGGTGAATGGAACCTTGAAT 300
Db AlathrglyshySerGlnlylethrglnSerValysanglyvalandlythleuVal 849

QY 301 GGTAAATGGGTTATCTCAAGCAGACCAACTCTTCTPAAAACTTTTGGACATCAAG 360
 DB 850 GYAASGlyLeuSerGlnAlaGlnAlaThrThrLeuSerLysAsnPheserApIleLys 869
 QY 361 AAGAGTTGAATGCAAACTTGAAATTTCAATACATTAACATTAATGACTCAAAAC 420
 DB 870 LysGlnLeuAsnAlaLysLeuGlyAsnPhesAsnAsnAsnAsnAsnAsnGlnGlyLeuLysAsn 889
 QY 421 GAACCATTTAATGCTAAAGTATATAAAGAAAGCGGGCAGACGACTGACCTTGAGAA 480
 DB 890 GlnProIleTyraLysValaLysValaLysValaGlyGlnAlaLaserLeuGlnGlu 909
 QY 481 CCCATTACGCTCAAGTGTCTAAAGGTAATGCAAAATGACCGACTCAATCAATA 540
 DB 910 ProIleTyraLysGlnAlaAlaLysLysValaAsnAlaLysLysLeuAsnGlnIle 929
 QY 541 GCAAGTGGTTGGTGTGTGAGGCAAGCAGCGGCTCCCTTTGAAAAGGCATGATAA 600
 DB 930 AlaSerGlyLeuGlyValaValaGlyGlnAlaLaglyPheProLeuLysArgHisAspLys 949
 QY 601 GTTGATGATCTCAGTAAGGAGGCTTTCAAGAAATCAAGATGGCTCAAAAATTGAC 660
 DB 950 ValAspAspLeuSerLysValaGlyLeuSerArgAsnGlnGlnLeuAlaGlnLysIleAsp 969
 QY 661 AATCTCAATCAAGCGGTATCAGAA 684
 DB 970 AsnLeuAsnGlnAlaValaSerGln 977

RESULT 3

Q6VR05_HELPEY PRELIMINARY; PRT; 1198 AA.
 AC Q6VR05_

DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE Caga.
 GN Name=caga; ORFNames=HP0547;
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 NCBI_TaxId=210;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Ca52;
 RX PubMed=15019987; DOI=10.1016/j.gene.2003.11.029;
 RA Blomestegren A., Lundin A., Nilsson C., Engstrand L., Lundberg J.;
 RT "Comparative analysis of the complete cag pathogenicity island
 sequence in four Helicobacter pylori isolates."
 RL Gene 328:85-93(2004)
 DR EMBL; AY330637; AAR03881.1; -; Genomic_DNA.
 DR GO; GO:0019534; F:toxin transporter activity; IRA.
 DR InterPro; IPR005169; Caga.
 DR InterPro; IPR004355; IVSec_caga.
 DR Pfam; PF03507; Caga; 1.
 DR PRINTS; PRO1553; TYPE4SSCAGA.
 SQ SEQUENCE 1198 AA; 133091 MW; DBBFSB554663BCOE CRC64;

Alignment Scores:

Pred. No.: 6.22e-66 Length: 1198
 Score: 1035.00 Matches: 214
 Percent Similarity: 83.2% Conservative: 4
 Best Local Similarity: 81.7% Mismatches: 10
 Query Match: 88.4% Indels: 34
 DB: 2 Gaps: 1

US-09-360-685C-26 (1-685) x Q6VR05_HELPEY (1-1198)

QY 1 AAAAAATGCAAAATTAAGATTTTCAGCAAGTAACGACGCAAGAAAAAGCGACTTGAAAAAT 60
 DB 766 LysAsnGlyLysAsnLysAspPheSerLysValThrGlnAlaLysSerAspLeuGlnLys 785
 QY 61 TCCGTTAAAGATGATATCATCAATCAAAAGTAACGATTAAGTTGATATCTCAATCA 120

DB 786 SerIleLysAspValIleIleAsnGlnLysIleThrAspLysValaAspAsnLeuAsnGln 805
 QY 121 GCGGTATCAGTGGCTTAAAGCAACGGGTGATTTCACTAGCTGTAAGCAACGCTTACCCGAT 180
 DB 806 AlaValSerIleAlaLysAlaThrGlyAspPheSerGlyValaGlnGlnAlaLeuAlaAsp 825
 QY 181 CTCAAAAATTTTCAAGAGCAATTGGCCCAACAGCTCAAAAAATGAAGTCTCAAT 240
 DB 826 LeuLysAsnPheserLysGlyGlnLeuAlaGlnGlnAlaGlnLysAsnGlnLysAsn 845
 QY 241 GCTAGCAAAAAATCGAAATATATCAATCCGTTAAGATGATGTAAGTAAGAACCTTACTC 300
 DB 846 ThrGlyLysAsnSerGlnIleTyraLysValaLysValaAsnGlyValaGlnIleVal 865
 QY 301 GGTAAATGGGTTATCTCAAGCAGACCAACTCTTCTPAAAACTTTTGGACATCAAG 360
 DB 866 GYAASGlyLeuSerLysAlaGlnAlaThrThrLeuSerLysAsnPheserApIleLys 885
 QY 361 AAGAGTTGAATGCAAACTTGAAATTTCAATACATTAACATTAATGACTCAAAAC 420
 DB 886 LysGlnLeuGlnAlaLysLeuGlyAsnPhesAsnAsnAsnAsnAsnAsnGlyLeuLysAsn 905
 QY 421 GAACCATTTAATGCTAAAGTATATAAAGAAAGCGGGCAGACGACTGACCTTGAGAA 480
 DB 906 GlnProIleTyraLysGlnAlaAlaLysLysValaAsnAlaLysLysLeuAsnGlnIle 925
 QY 481 CCCATTACGCTCAAGTGTCTAAAGGTAATGCAAAATGACCGACTCAATCAATA 540
 DB 926 ProIleTyraLysGlnAlaAlaLysLysValaAsnAlaLysLysLeuAsnGlnIle 945
 QY 541 GCAAGTGGTTGGTGTGTGAGGCAAGCAGCGGCTCCCTTTGAAAAGGCATGATAA 576
 DB 946 AlaSerGlyLeuGlyValaGlyGlnAlaLaglyPheProLeuLysArgHisAspLys 965
 QY 576 ----- 576
 DB 966 ValAspAspLeuSerLysValaGlyLeuSerAlaAsnHisGlnProIleTyraLysThrIle 985
 QY 577 -----TTCCCTTGAAGCAAGCATGATAAGTTGATGATCTCAGTAAG 618
 DB 986 AspAspLeuGlyGlySerPheProLeuLysArgHisAspLysValaAspAspLeuSerLys 1005
 QY 619 GTAGGCTTTTCAGCAAGATCAAGAAATGGCTCAGAAAATTGACATCTCAATCAAGCGGTA 678
 DB 1006 ValGlyLeuSerArgAsnGlnGlnLeuAlaGlnLysIleAspAsnLeuAsnGlnAlaVal 1025
 QY 679 TCAGAA 684
 DB 1026 SerGln 1027

RESULT 4

Q6VRM7_HELPEY PRELIMINARY; PRT; 1181 AA.
 AC Q6VRM7_

DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE Caga.
 GN Name=caga; ORFNames=HP0547;
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 NCBI_TaxId=210;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Ca73;
 RX PubMed=15019987; DOI=10.1016/j.gene.2003.11.029;
 RA Blomestegren A., Lundin A., Nilsson C., Engstrand L., Lundberg J.;
 RT "Comparative analysis of the complete cag pathogenicity island
 sequence in four Helicobacter pylori isolates."
 RL Gene 328:85-93(2004)
 DR EMBL; AY330639; AAR03909.1; -; Genomic_DNA.

DR GO: GO:0019534; F:toxin transporter activity; IEA.
DR InterPro: IPR005169; Caga.
DR InterPro: IPR004355; IYSec_caga.
DR Pfam: PF03507; Caga_1
DR PRINTS: PRO1553; TYPE4SCGA.
SEQUENCE 1181 AA; 131427 MW; A991CAF617CDE3E CRC64;

Alignment Scores:

Pred. No.:	1,686-65	Length:	1181
Score:	1029.00	Matches:	213
Percent Similarity:	82.8%	Conservative:	4
Best Local Similarity:	81.3%	Mismatches:	11
Query Match:	87.9%	Indels:	34
DB:	2	Gaps:	1

US-09-360-685C-26 (1-685) x OGVN7_HELPY (1-1181)

```
QY 1 AAAATGCGCAAAATTAAGGATTTTCAGCAGGTAAAGCAAGCAAAAGCGACTTGAAAT 60
Db 750 LysanGlyLysanLysanLysanPheSerLysValThrGlnAlaLysSerPheLysLysan 769
QY 61 TCCGTTAAAGATGTGATCATCAATCAAAAGTAAACGGATTAAGTTGATCAATCAATCA 120
Db 770 SerLysPheValIleIleLysGlnLysIleThrPheLysValLysanLysSer 789
QY 121 GCGGTATCAGTGTGATTAAGCAAGCGGTGATTTTCAGTGGTAAAGCAAGCGTTAGCCGAT 180
Db 790 AlaValSerValAlaLysAlaThrGlyPhePheSerLysValGlnGlnValLysAlaGly 809
QY 181 CTCAAAAATTTCTCAAGAGCAATGGCGCCCAAGCAAGCTCAAAAAATGAATCTCAAT 240
Db 810 LeuLysanPheSerLysGlnGlnLysanLysGlnLysanLysanLysanLysan 829
QY 241 GCTGAAAAAATCTGAAATTAATCAATCCGTTAAAGATGGTGAATGAACCTTATGC 300
Db 830 ThrGlyLysLysSerGlnLysIleThrLysValLysanGlyValLysanGlyThrLysVal 849
QY 301 GGTATGGGTTATCTCAAGAGCAAGCAAGCTTTTCTTAAAGCTTTTGGACATCAAG 360
Db 850 GlyAnGlyLysSerGlnLysIleThrLysSerLysanPheSerLysIleLys 869
QY 361 AAAGAGTTGAATGCAAACTTGAATTTCAATTAACAATTAATGAATGAATGAATGA 420
Db 870 LysGlnLysanLysLysLysanLysanPheLysanLysanLysanLysanLysan 889
QY 421 GAACCATTTATGCTTAAGTTAAAGTTAAAGAAAGCAAGGCAAGCAAGCTTGAAGAA 480
Db 890 GlnPheLysValLysValLysanLysLysValLysValLysValLysValLysVal 909
QY 481 CCATTTAGCTCAAGTTGCTTAAAGGTTAAAGTAAATGCAAAATTTGACCACTCAATA 540
Db 910 ProLysValLysanLysValLysValLysValLysValLysValLysValLysVal 929
QY 541 GCAAGTGGTTGGTGTGTAGGCGCAAGCGGCGC----- 576
Db 930 AlaSerGlyLysGlyValGlyGlnAlaLysGlyPheProLysLysIleLysLys 949
QY 576 ----- 576
Db 950 ValAspPheLysSerLysValGlyLysSerAlaAsnProGlnPheLysValLysIle 969
QY 577 -----TTCCCTTTGAAAAGGCAATGATTAAGTTAGTAAATGCTCAAG 618
Db 970 AspGlnLysanLysPheProLysLysValLysIleLysValLysanLysSerLys 989
QY 619 GTAGGCTTTCAAGAAATCAAGAAATGCTCAAGAAATTTGCAATCAATCAAGCGGTA 678
Db 990 ValGlyLysSerLysanGlnLysLysLysLysLysLysLysLysLysLysLysLys 1009
QY 679 TCAGAA 684
Db 1010 SerGln 1011
```

RESULT 5
OAPL5_HELPY
ID OAPL5_HELPY PRELIMINARY; PRT; 1230 AA.

AC OAPL5;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
CN Cytoxin-associated protein A.
OS Name-caga;
OC Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
KM (1)
KP NOCLEOTIDE SEQUENCE.
KC STRAIN=7bgs;
KA Kim S.Y., Blaser M.J., Lee Y.C., Pilling M.H.;
KT Helicobacter pylori stimulates matrix metalloproteinase-1 secretion
RT from gastric epithelial cells via Caga-dependent and independent
RT mechanisms: requirement for Erk activation."
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; D0067454; AA568033.1; -; Genomic DNA.
SQ SEQUENCE 1230 AA; 136662 MW; 648240E8B3001351 CRC64;

Alignment Scores:

Pred. No.:	1,676-65	Length:	1230
Score:	1029.00	Matches:	218
Percent Similarity:	70.6%	Conservative:	5
Best Local Similarity:	69.0%	Mismatches:	8
Query Match:	87.9%	Indels:	88
DB:	2	Gaps:	1

US-09-360-685C-26 (1-685) x OAPL5_HELPY (1-1230)

```
QY 1 AAAATGCGCAAAATTAAGGATTTTCAGCAGGTAAAGCAAGCAAAAGCGACTTGAAAT 60
Db 746 LysanGlyLysanLysanLysanPheSerLysValThrGlnAlaLysSerPheLysLysan 765
QY 61 TCCGTTAAAGATGTGATCATCAATCAAAAGTAAACGGATTAAGTTGATCAATCAATCA 120
Db 766 SerLysPheValIleIleLysGlnLysIleThrPheLysValLysanLysanLysan 785
QY 121 GCGGTATGCTTAAAGCAAGCGGTGATTTTCAGTGGTAAAGCAAGCGTTAGCCGAT 180
Db 786 AlaValSerValAlaLysAlaThrGlyPhePheSerLysValGlnGlnValLysAlaAsp 805
QY 181 CTCAAAAATTTCTCAAGAGCAATGGCGCCCAAGCAAGCTCAAAAAATGAATGTCAT 240
Db 806 LeuLysanPheSerLysGlnGlnLysanLysGlnGlnLysanLysanLysanLysan 825
QY 241 GCTGAAAAAATCTGAAATTAATCAATCCGTTAAAGATGGTGAATGAACCTTATGC 300
Db 826 ValGlyLysLysSerGlnLysIleThrLysValLysanGlyValLysanGlyThrLysVal 845
QY 301 GGTATGGTATCTCAAGAGCAAGCAAGCTTTTCTTAAAGCTTTTGGACATCAAG 360
Db 846 GlyanGlyLysSerLysLysIleThrLysSerLysPheSerLysIleLys 865
QY 361 AAAGAGTTGAATGCAAACTTGAATTTCAATTAAGTAAATGCAATTAATGATCAAAAAAC 420
Db 866 LysGlnLysanLysLysLysanLysanPheLysanLysanLysanLysanLysan 885
QY 421 GAACCATTTATGCTTAAAGTTAAAGTTAAAGAAAGCAAGGCAAGCAAGCTTGAAGAA 480
Db 886 GlnPheLysValLysValLysanLysLysValLysValLysValLysValLysVal 905
QY 481 CCATTTAGCTCAAGTTGCTTAAAGGTTAAAGTAAATGCAAAATTTGACCACTCAATA 540
Db 906 ProLysValLysanLysValLysValLysValLysValLysValLysValLysVal 925
QY 541 GCAAGTGGTTGGTGTGTAGGCGCAAGCGGCGC----- 576
Db 926 AlaSerGlyLysGlyValGlyGlnAlaLysGlyPheProLysLysIleLysLys 945
```



```

QY 576 ----- 576
Db 946 ValAspAspLeuSerIysValGIArgSerValSerProGluProIleTyrAlaGlnVal 965
QY 576 ----- 576
Db 966 AlaIylLeuValAsnAlaIylSileAspArgLeuAsnGlnIleAlaSerGIylLeuGIylGly 985
QY 576 ----- 576
Db 986 ValGIylGlnAlaIlaGlyPheProLeuIylsArgHisAspIylsValAspAspLeuSerIys 1005
QY 576 ----- 576
Db 1006 ValGIylArgSerValSerProGluProIleTyrAlaThrIleAspAspLeuGIylGIylPro 1025
QY 577 TTCCTTTGAAAAAGCATGATAAGTTGATGATCTCATAGGTAAGGCTTTTCAGAGAA 636
Db 1026 PheProLeuIylsArgHisAspIylsValAspAspLeuSerIysValGIylLeuSerArgAsn 1045
QY 637 CAGAAATTGGCTCAGAAAAATTGCATCTCATCAACGGGTATCAGAA 684
Db 1046 GlnIylsLeuAlaGlnIylsIleAspAsnLeuAsnGlnAlaValSerGln 1061

```

RESULT 6

O8RNUO_HELPY

O8RNUO_HELPY PRELIMINARY; PRT; 267 AA.

```

AC O8RNUO;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytotoxin-associated antigen (Fragment).
GN Name-caga;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxId=210;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CH-CTX1;
RX MEDLINE=20123532; PubMed=10660136;
RA Opazo P., Muller I., Rollan A., Valenzuela P., Yudelevich A.,
RA Garcia-de la Guardia R., Urra S., Venegas A.;
RT "Serological response to Helicobacter pylori recombinant antigens in
RT Chilean infected patients with duodenal ulcer, non-ulcer dyspepsia and
RT gastric cancer.";
RL APMIS 107:1069-1078 (1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CH-CTX1;
RX MEDLINE=22121498; PubMed=12125208;
RA Muller I., Medina-Selby A., Palacios J.L., Martinez P., Opazo P.,
RA Bruce E., Mancilla M., Valenzuela P., Yudelevich A., Venegas A.;
RT "Cloning and comparison of ten gene sequences of a Chilean H. pylori
RT strain with other H. pylori strains revealed higher variability for
RT vacA and caga virulence factors.";
RL Biol. Res. 35:67-84 (2002).
DR EMBL; AF479032; AAL86902.1; -; Genomic_DNA.
DR InterPro; IPR005169; Caga.
DR Pfam; PF03507; Caga; 1.
FT NON_TER 1
FT NON_TER 267
SQ SEQUENCE 267 AA; 29032 MW; 04B4F514809608B8 CRC64;

```

Alignment Scores:

```

Pred. No.: 3,3e-65 Length: 267
Score: 1025.50 Matches: 216
Percent Similarity: 83.4% Conservative: 5
Best Local Similarity: 81.5% Mismatches: 7
Query Match: 87.6% Indels: 37
DB: 2 Gaps: 3

```

US-09-360-685C-26 (1-685) x O8RNUO_HELPY (1-267)

```

QY 1 AAAATGGCAAAATATAGATTTCAGCAAGGTATAGCAAGCAAAAAGCACTTGAAAT 60
Db 3 LysAsnGIylLysAsnIylsAspPheSerIysValThrGlnAlaIylSerAspLeuGln 22
QY 61 TCCGTTAAAGATGTGATCATCAATCAAAAGTAAGGATTAAGTTGATTAATCTCAATCA 120
Db 23 SerIleLysAspValIlePheAsnGlnIylsIleThrAspIylsValAspAspLeuAsn 42
QY 121 GCGGTATCAGTGGCTTAAGCAACGGGTGATTTCAGTAGGGTACAGCAAGCTTAGCCGAT 180
Db 43 AlaValSerValAlaIylsAlaThrGIylsAspPheSerArgValGIylGlnIleAlaAsp 62
QY 181 CTCAAAAATTTCTCAAGAGCAATTTGCCCAACAGCTCAAAAAATGAAGTCTCAAT 240
Db 63 LeuIylsAsnPheSerIylsGIylGlnIleAlaGlnIylsAlaGlnIylsAsnIylSerLeuAsn 82
QY 241 GGTAGAAAAAATCTGAAATATATCATATCCCTTAAGAAATGCTGAATGGAACCTAGTC 300
Db 83 AlaGIylLysSerGIylIleTyrAlaIylsValAsnGIylValSerGIylThrLeuVal 102
QY 301 GGTAAATGGTATCTCAAGCAGACCAACCACTCTTCTTAAAACTTTTCGACATCAAG 360
Db 103 GIylsAsnGIylLeuSerGIylsAlaGlnIylsAlaThrThrLeuSerIylsAsnPheSerIyls 122
QY 361 AAAGATTGATGCAAAACTT---GGAAATTTCAATTAACATTAACATTAATGACTCAAA 417
Db 123 LysGIylLeuAsnAlaIylsLeuPheGIylsAsnPheAsnAsnAsnAsnGIylsLeu 142
QY 418 AAC-----GAACCACTTATCTTAAGTTAATTAATAAAGCAAGCGGACAGCTTAC 471
Db 143 AsnSerThrGlnProIleTyrAlaIylsValAsnIylsGIylGlnIleAlaSer 162
QY 472 CTTAGAAGAACCATTAAGCTCAAGTGTCTAAAGGTAAAGGTAAAGCAAAATTAACGACTC 531
Db 163 ProGluGluProIleTyrThrGlnValAlaIylsValAsnAlaIylsIleAspArgLeu 182
QY 532 AATCAAAATGCAAGTGGTGTGGGTGTAGGGCAAGCAGCGGC----- 576
Db 183 AsnGlnIleAlaSerGIylLeuGIylValGIylsAlaIylsPheProLeuIylsArg 202
QY 576 ----- 576
Db 203 HisAspIylsValAspAspLeuSerIylsValGIylArgSerValSerProGluProIleTyr 222
QY 577 ----- 577
Db 223 AlaThrIleAspAspLeuGIylGIylProPheProLeuIylsArgHisAspIylsValAsp 242
QY 610 CTCAGTAAGTAGGCTTTCAAGATCAAGATTCAGAAATTCAGAAATTCAGATTCAT 669
Db 243 LeuSerIylsValGIylLeuSerArgHisAsnGlnIleuAlaGlnIylsIleAspAsnLeuAsn 262
QY 670 CAGCGGTATCAGAA 684
Db 263 GlnAlaValSerGln 267

```

RESULT 7

O52GZ6_HELPY

O52GZ6_HELPY PRELIMINARY; PRT; 1142 AA.

```

AC O52GZ6;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Caga (Fragment).
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxId=210;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2808;

```


QY	361	AAAGGTGGAAGCAAACTTGGAAATTCATTAACAATTAATGACTCAAAAAC	420
Dd	870	LysGIuLeuAenAlaIysLeuGIyAsnPhenAenAsnAsnAsnAsnIyLeuIyAsn	889
QY	421	GAACCCATTATATGCTAAAGTTAATAAAAAAAGAACGAGGCAAGACTAGCCTTGAGNA	480
Dd	880	GIuPProIleIyAlaIyAlaValAsnIySuIySesGIyGlnAlaIleSerProGIuGln	909
QY	481	CCCATTTAGCGCTCAAGTGTCTAAAAAGTTAAATGCAAAAATTGACCGACTCATCAATA	540
Dd	910	ProIleIyAlaGlnValaIaIySuIyValaIaAlaIyAlaIleAspArgLeuAsnGlnIle	929
QY	541	GCAAGTGTGTTGGGTGTTGTGAGGCAAGCAGCGGGC-----	576
Dd	930	AlaSerGIyLeuGIyGIyValGIyGlnAlaAlaGlyPheProLeuIyArgHisAspIyS	949
QY	576	-----	576
Dd	950	ValAspAspLeuSerIyValGIyArgSerValSerProGIuProIleIyAlaThrIle	969
QY	577	-----TTCCCTTGAAAGAGCAGTAAAGTTGATGATGATTCAGTAG	618
Dd	970	AspAspLeuGIyGlyProPheProLeuIyAsnGHisAspIySValaAspAspLeuSerIyS	989
QY	619	GTAAGGCTTTCAAGGAATCAAGAAATTTGGCTCAGAAATTTGACAACTCATCAACCGGTA	678
Dd	990	ValGIyLeuSerArgAsnGlnIuLeuAlaGlnIySileAspSerLeuAsnGlnAlaVal	1009
QY	679	TCAGAA 684	
Dd	1010	SerIyS 1011	

	RESULT	9
ID	Q7X4U1_HELPY	PRT; 326 AA.
AC	Q7X4U1;	
DT	01-OCT-2003 (TrEMBLrel. 25, Created)	
DR	01-OCT-2003 (TrEMBLrel. 25, Last sequence update)	
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)	
DE	Caga (Fragment).	
GN	Name=Caga;	
OS	Helicobacter pylori (Campylobacter pylori).	
OC	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;	
CC	Helicobacteraceae; Helicobacter.	
CK	NCBI_TaxID=210;	
RN	[1]	
RP	NUCLEOTIDE SEQUENCE.	
RC	STRAIN=China:079a;	
XX	PubMed=14734191; DOI=10.1016/S0928-8244(03)00299-2;	
RA	Zhou W., Yamazaki S., Yanakawa A., Ohtani M., Ito Y., Keida Y.,	
RT	Higashi H., Hatakeyama M., Si J., Azuma T.;	
RT	"The diversity of vacA and cagA genes of Helicobacter pylori in East Asia";	
RL	FEMS Immunol. Med. Microbiol. 40:81-87(2004).	
DR	EMBL; AB110563; BAC77006.1; -; Genomic_DNA.	
DR	InterPro; IPR005169; Caga.	
DR	Pfam; PF03507; Caga; 1.	
FT	NON_TER	1
FT	NON_TER	1
SO	SEQUENCE	326 AA; 35988 MW; 7BC601BFA98ACABF CRC64;

Alignment Scores:

Pred. No.:	1,136-63	Length:	1,326
Score:	90.00	Matches:	5
Percent Similarity:	90.4%	Conservative:	201
Best Local Similarity:	88.2%	Mismatches:	22
Query Match:	85.7%	Indels:	85
DB:	2	Gaps:	0

US-09-360-685C-26 (1-685) x Q7X4J1_HELPY (1-326)

QY 1 AAAAATGGCAAAATAAGATTTCAGCAAGGTAACGCAAGCAAAAAGCGACCTTGAAAT 60

[illegible]

RESULT 10

ID Q9L5X9 HELPY PRELIMINARY; PRT; 1179 AA.

AC Q9L5
DT 01-0

DT 01-OCT-2000 (Tremblay, 15, last sequence update)

DT 01-JUN-2003 (Tremblay, 24, Last annotation update)

DE cytolexin-associated protein caga.
Name-caga:

OS Helicobacter pylori (Campylobacter pylori).

OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;

OC Helicobacteraceae; Helicobacter.
 OY NCBI TaxID=310.
 OX

ONCE - [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=SS1;
DA zbase T F

Submitted (APR-2000) to the EMBL/GenBank/DBJ databases

RN [2]

RP NUCLEOTIDE

RC STRAIN=SS1; Zhang J. Z.:
RA Zhou J. C.

Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF247651; AAF63759.1; -; Genomic DNA.

DR GU; GU:0019534; F:toxIn transporter activity; IEA

SECRET

DR InterPro: IPR004355; IYSec_caga.
DR Pfam: PF03507; Caga; 2.
DR PRINTS: PRO1553; TYPE4SSCAGA.
SQ SEQUENCE 1179 AA; 131297 MW; 79A112P58B749787 CRC64;

Alignment Scores:

Pred. No.:	1,12e-63	Length:	1179
Score:	1003.50	Matches:	209
Percent Similarity:	81.7%	Conservative:	5
Best Local Similarity:	79.8%	Mismatches:	13
Query Match:	85.7%	Indels:	35
DB:	2	Gaps:	2

US-09-360-685C-26 (1-685) x Q9L5X9_HELPY (1-1179)

```
QY 1 AAAAATGCAAAAATTAAGATTTCACGAGGTAAGCAAGCAAAAAGCACTTGAAAT 60
Db 750 LysAsnGlyLysAsnLysAspPheSerValThrGlnAlaLysSerAspLeuGlnAsn 769
QY 61 TCGTTAAAGATGATCATCAATCAAAAGTAAGCAATTAAGTGAATCTCAATCAA 120
Db 770 SerValLysAspValIleIleAsnGlnLysIleThrAspLysValAspAsnLeuAsnGln 789
QY 121 GCGGTATCGTGGCTAAAGCAACGGGTGATTCAGTGGGTAGACCAAGCGTTAGCCGAT 180
Db 790 AlaValSerValAlaLysAlaThrGlnLysAspPheSerValGlnGlnAlaLeuAlaAsp 809
QY 181 CTCAAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCAAAAATGAAGTCAAT 240
Db 810 LeuLysAsnPheSerLysGlnGlnLeuAlaGlnGlnAlaGlnLysAsnGlnAspPheAsn 829
QY 241 GCTAGAAAAAATCTGAATATATCATCCGTTAAGATGGTGTAGTGAACCTTAGTC 300
Db 830 ThrGlyLysAsnSerAlaLeuTyrgInSerValLysAsnGlyValAsnGlyThrLeuVal 849
QY 301 GGTATGGGTATCTCAAGCAAGCAACCACTTTCTTAAAACTTTGGACATCAAG 360
Db 850 GlyAsnGlyLeuSerLysAlaGlnAlaThrThrLeuSerLysAsnPheSerAspIleLys 869
QY 361 AAAGATGATGCAAAACTTGAAATTTCAATTAACATTAACATTAATGATCAAAAAC 420
Db 870 LysGlnLeuAsnAlaLysLysGlnLysAsnPheAsnAsnAsnAsnAsnGlyLeuLysAsn 889
QY 421 GAACCATTTAGCTAAAGTTAATTAATAAAGCAAGCAAGCAAGCTTAGAAGA 480
Db 890 GluProIleTyraLysGlnValAsnLysLysLysAlaGlnValAlaSerProGlnGln 909
QY 481 CCAATTACGCTCAAGTTGCTAAAGGTAATTCGAAAATTTGACCCGATCAATCAATA 540
Db 910 ProIleTyraLysGlnValAlaLysLysValAsnAlaLysLysAspGlnLeuAsnGlnAla 929
QY 541 GCAAGTGGTGGGTGTTGTAAGGCAAGCAGCGGCTTCCTTTGAAAAGCATGATAA 600
Db 930 AlaSerGlyLeuGlyLysGlyLysGln--AlaGlyPheProLeuLysAsnGlnAspLys 948
QY 601 GTTGAT----- 606
Db 949 ValAspAspLeuSerLysValGlyArgSerValSerProGluProIleTyraLThrIle 968
QY 607 -----GATCTCAGTAG 618
Db 969 AspAspLeuGlyGlyProPheProLeuLysLysIleAlaLysValGlnAspLeuSerLys 988
QY 619 GTAGGGCTTTCAGGAATCAAGATTTGGCTCAAGAAAATGCAATCTCAATCAAGCGGA 678
Db 989 ValGlyLeuSerArgGlnGlnGlnLeuThrGlnLysLysIleAspAsnLeuAsnGlnAlaVal 1008
QY 679 TCAGAA 684
Db 1009 SerGln 1010
```

RESULT 11
Q52GZ5_HELPY

ID Q52GZ5_HELPY PRELIMINARY; PRT, 1186 AA.
AC Q52GZ5;
DT 13-SEP-2005 (Tremblrel. 31, Created)
DT 13-SEP-2005 (Tremblrel. 31, Last sequence update)
DT 13-SEP-2005 (Tremblrel. 31, Last annotation update)
DE Caga (Fragment).
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=P310;
RX PubMed=15972330; DOI=10.1073/pnas.0409873102;
RA Brandt S., Kwok T., Hartig R., Konig W., Backert S.;
RT "NF-(kappa)B activation and potentiation of proinflammatory responses
by the Helicobacter pylori Caga protein".
RL Proc. Natl. Acad. Sci. U.S.A. 102:9300-9305 (2005).
DR EMBL: DQ011620; RefSeq: NP_053961.1; Genomic DNA.
DR GO: GO:0019554; F:toxin transporter activity; IEA.
DR InterPro: IPR005169; Caga.
DR InterPro: IPR004355; IYSec_caga.
DR Pfam: PF03507; Caga; 1.
DR PRINTS: PRO1553; TYPE4SSCAGA.
FT NON_TER 1186
SQ SEQUENCE 1186 AA; 131865 MW; CD5AC9D1DA41BF3F CRC64;

Alignment Scores:

Pred. No.:	7.47e-63	Length:	1186
Score:	992.00	Matches:	212
Percent Similarity:	80.7%	Conservative:	5
Best Local Similarity:	78.8%	Mismatches:	10
Query Match:	84.7%	Indels:	42
DB:	2	Gaps:	3

US-09-360-685C-26 (1-685) x Q52GZ5_HELPY (1-1186)

```
QY 1 AAAAATGCAAAAATTAAGATTTCACGAGGTAAGCAAGCAAAAAGCACTTGAAAT 60
Db 750 LysAsnGlyLysAsnLysAspPheSerValThrGlnAlaLysSerAspLeuGlnAsn 769
QY 61 TCGTTAAAGATGATCATCAATCAAAAGTAAGCAATTAAGTGAATCTCAATCAA 120
Db 770 SerValLysAspValIleIleAsnGlnLysIleThrAspLysValAspAsnLeuAsnGln 789
QY 121 GCGGTATCGTGGCTAAAGCAACGGGTGATTCAGTGGGTAGACCAAGCGTTAGCCGAT 180
Db 790 AlaValSerValAlaLysAlaThrGlnLysAspPheSerValGlnGlnAlaLeuAlaAsp 809
QY 181 CTCAAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCAAAAATGAAGTCAAT 240
Db 810 LeuLysAsnPheSerArgGlnGlnLeuAlaGlnGlnAlaGlnLysAsnGlnSerLeuAsn 829
QY 241 GCTAGAAAAAATCTGAATATATCATCCGTTAAGATGGTGTAGTGAACCTTAGTC 300
Db 830 AlaGlyLysLysSerGlnLysIleTyrgInSerValLysAsnGlyValAsnGlyThrLeuVal 849
QY 810 GTTATGGGTATCTCAAGCAAGCAACCACTTTCTTAAAACTTTGGACATCAAG 360
Db 850 GlyAsnGlyLeuSerGlyIleGlnAlaThrAlaLeuAlaLysAsnPheSerAspIleLys 869
QY 361 AAAGATGATGCAAAACTTGAAATTTCAATTAACATTAACATTAATGATCAAAAAC 420
Db 870 LysGlnLeuAsnGlnLysLysLysAsnLysLysLysAlaGlnValAlaSerProGlnGln 909
QY 421 -----GATCTCAGTAG 618
Db 889 GlyLysAspLysGlyProGlnGlnProIleTyraLysValAsnLysLysLysThrGly 908
QY 460 CAAGCAGTAGCTTGAAGAACCACTTAGCGTCAAGTTGAAAAGTGAATGCAAAA 519
Db 909 GlnAlaLysSerProGlnGlnProIleTyraLysGlnValAlaLysLysValAsnAlaLys 928
```

```

QY 520 ATTGACGACTCATCAATCAATAGCAAGTGTGTTGGTTAGGCGCAAGCAGCGGC--- 576
DB 929 ILeaspArgLeuSnmLysIleAlaSerGlyLeuGlyGlyValGlyGlnAlaAlaGlyPhe 948
QY 576 ----- 576
DB 949 ProLeuLysLysHisAspLysValAspAspLeuSerLysValGlyArgSerValSerPro 968
QY 577 -----TTCCCTTTGAAAAGCAATGAT 597
DB 969 GluProIleTyrAlaThrIleAspAspLeuGlyGlyProPheProLeuLysLysHisAsp 988
QY 598 AAGTTGATGATCTCATGTAAGTAGGCTTTCAAGCAATCAAGATTTGGCTCAGAAATT 657
DB 989 LysValAspAspLeuSerLysValGlyLeuSerArgAsnGlnLysLeuAlaGlnLysIle 1008
QY 658 GACAACTTCATCAACGCGGTATCAGAA 684
DB 1009 AspAsnLeuAsnGlnAlaValSerGlu 1017

RESULT 12
Q9F222_HELPY PRELIMINARY; PRT; 1179 AA.
ID Q9F222_HELPY PRELIMINARY; PRT; 1179 AA.
AC Q9F222;
DC 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Caga.
GN Name=caga;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_Taxid=210;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC43579;
RX MEDLINE=21457648; PubMed=11573724; DOI=10.1007/s005300700002;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RA Ando T.;
RT "Heterogeneity found in the caga gene of Helicobacter pylori from
RT Japanese and non-Japanese isolates.";
RL J. Gastroenterol. 35:890-897(2000).
DR EMBL; AB015414; BAB20924.1; -; Genomic DNA.
DR GO; GO:0019534; F:toxin transporter activity; IEA.
DR InterPro; IPR005169; Caga.
DR InterPro; IPR004355; IVSec_caga.
DR Pfam; PF03507; Caga; 1.
DR PRINTS; PR01553; TYPE4S8CAGA.
SQ SEQUENCE 1179 AA; 131686 MW; 7D0ADC34B6446434 CRC64;

Alignment Scores:
Pred. No.: 9,57e-63 Length: 1179
Score: 990.50 Matches: 206
Percent Similarity: 80.8% Conservative: 5
Best Local Similarity: 78.9% Mismatches: 17
Query Match: 84.6% Indels: 33
DB: 2 Gaps: 1

US-09-360-685C-26 (1-685) x Q9F222_HELPY (1-1179)
QY 1 AAAAAAGCAAAATAAGATTTCAGCAGTAAAGCAAGCAAAAGCAAGCAAGCAAGCAAGT 60
DB 750 LysValAsnGlyLysAsnLysAspPheSerLysValThrGlnAlaLysSerAspLeuGln 769
QY 61 TCCGTTAAAGATGTGATCATCAATCAAAAGTAAAGCATTAAGTATCATCATCAATCAA 120
DB 770 SerValLysAspAlaIlePheAsnGlnLysValThrAspLysValAspAspLeuAsnGln 789
QY 121 GCGGTATCATGTGCTTAAGCAACGCGGTGATTTCAAGGCTTAAGCAACGCTTACCGCAT 180
DB 790 AlaValSerValAlaLysAlaThrSerAspPheSerArgValGlnGlnAlaLeuAlaAsp 809
QY 181 CTCAAAAATTCTCAAGAGGACATTTGGCCCAACAAAGCTCAAAAAATAAAGTCTCAAT 240

```

```

DB 810 LeuLysAsnPheSerLysGlnLysLeuAlaGlnGlnThrGlnLysAsnLysSerPheAsn 829
QY 241 GCTAGAAAAAATCTGAATATATATCAATCCGTTAAGATGTGTGAATGAACCCCTAGTC 300
DB 830 ValGlyLysLysSerGlnLysIleTyrGlnSerValLysAsnGlyValAsnGlyThrLeuVal 849
QY 301 GGTATGGTTATCTCAAGCAGACAGACCACTCTTTCTTAAATAAATCTTTGGACATCAAG 360
DB 850 GlyThrGlyLeuSerLysAlaGlnAlaThrThrLeuSerLysAsnPheSerAspLysIleLys 869
QY 361 AAGAGTTGAATGCAAACTTGGAAATTCAATTAACAATTAACAATTAAGTCAAAAC 420
DB 870 LysGlnLeuAsnAlaLysLeuGlnLysAsnPheAsnAsnAsnAsnLysLeuLysAsn 889
QY 421 GAACCATTTATGCTAATGTAATTAATAAAGAAAGCAAGGCAAGCAGCTACCTTGAGAA 480
DB 890 GluProIleTyrAlaGlnValAlaLysLysValLysValGlyGlnAlaLysSerHisGlnGlu 909
QY 481 CCCATTACGCTCAAGTTGCTTAATAAAGTAAATGCAAAATTAAGCAAGCACTCAATCAATA 540
DB 910 ProIleTyrThrGlnValAlaLysLysValAsnAlaLysIleAspGlnLeuAsnGlnAla 929
QY 541 GCAAGTGTGTTGGTGTGTTAGGCAAGCA----- 570
DB 930 AlaSerGlyLeuGlyGlyAlaAsnGlnAlaGlyPheProLeuLysArgHisAspLysVal 949
QY 570 ----- 570
DB 950 AspAspLeuSerLysValGlyLeuSerAlaAsnHisGluProIleTyrAlaThrIleAsp 969
QY 571 -----GCGGCGCTTCCCTTTGAAAAGCATGATTAAGTTGATGATCTCAGTAAGTA 621
DB 970 AspLeuGlyGlyProPheProLeuLysArgHisAspLysValAspAspLeuSerLysVal 989
QY 622 GGGCTTCAAGCAATCAAGAAATTTGGCTCAGAAATTAAGCAATCTCAATCAACCGGTATCA 681
DB 990 GlyLeuSerArgAspGlnGlnLysLeuAlaGlnLysIleAspAsnLeuAsnGlnAlaValSer 1009
QY 682 GAA 684
DB 1010 Glu 1010

RESULT 13
Q6VRG7_HELPY PRELIMINARY; PRT; 1182 AA.
ID Q6VRG7_HELPY PRELIMINARY; PRT; 1182 AA.
AC Q6VRG7;
DC 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Caga.
GN Name=caga; ORFNames=HP0547;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_Taxid=210;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=DUB23.2;
RX PubMed=15019987; DOI=10.1016/j.gene.2003.11.029;
RA Blomstergren A., Lundin A., Nilsson C., Engstrand L., Lundberg J.;
RT "Comparative analysis of the complete cag pathogenicity island
RT sequence in four Helicobacter pylori isolates.";
RL Gene 328:85-93(2004).
DR EMBL; AY310644; AAR03970.1; -; Genomic DNA.
DR GO; GO:0019534; F:toxin transporter activity; IEA.
DR InterPro; IPR005169; Caga.
DR InterPro; IPR004355; IVSec_caga.
DR Pfam; PF03507; Caga; 1.
DR PRINTS; PR01553; TYPE4S8CAGA.
SQ SEQUENCE 1182 AA; 131615 MW; BDEC536BBAF1E27 CRC64;

Alignment Scores:

```

Pred. No.: 2,796-62 Length: 1182
 Score: 994.00 Matches: 207
 Percent Similarity: 80.3% Conservative: 5
 Best Local Similarity: 78.4% Mismatches: 16
 Query Match: 84.0% Indels: 36
 DB: 2 Gaps: 2

US-09-360-685C-26 (1-685) x Q6VNG7_HELPY (1-1182)

QY	1	AAAAATGCAAAATTAAGATTTCAGCAAGGTAAACGACAAAGCAAAAGCACTTGAAAT	60
DB	750	LYSANGILYSAENLYSAEPHESERYVALTHRGINLALYSERAPLEUGLUBEN	769
QY	61	TCGGTTAAAGATGATCATCAATCAAAAGTAACGATTAAGTTGATTCATCAATCA	120
DB	770	SERVALYSAEPVALLEILEANGILYSAVALTHRAPLYEVALASPAENLEUANGIN	789
QY	121	GCGGTATCAGTGGCTAAAGCAAGGCTGATTCAGTGGGTAGACGACGCTTAGCCGAT	180
DB	790	ALAVASERILEALYSALEILEGIYSAEPHESERYVALGLUGINLALAEUALAEP	809
QY	181	CTCAAAAATTTCTCAAGAGCAATTGGCCCAACAGCTCAAAAATTAAGATCTCAAT	240
DB	810	LEULYSAENPHESERYSGILUGINLEUINALGINLALYSAENGLUBAPHEAEN	829
QY	241	GCTAGAAAAATCTGAAATATATCAATCCGTTAAAGATGCTGATTAAGCACTTAGTC	300
DB	830	THRGILYLYLSESGILUILEYRGINSERVALYSAENGLYVALAENGLYTHRLAEVAL	849
QY	301	GCTAATGGTTATCTCAAGCAAGCAACCACTTTCTTAAAACTTTTGACATCAAG	360
DB	850	GLYSAENGLYSESGILYILEGINLATHRLALEUINALYSAENPHESERAPILEY	869
QY	361	AAAGATGTGATGCAAACTTGAAATTTCAATTAACATTAACATTAATGACTCAAAAAC	420
DB	870	LYSGILUEUANGILYSPHELYSAENPHESERAPHEAENLALYSAENGLYLEUGLUBEN	889
QY	421	-----GAACCCATTTATGCTAAAGTTAATTAATAAAGAAAGCAAGGCAAGCACTT	474
DB	890	SETHRGILUPROILEYTRALALYSVALASNLYSLEYSTHGGILNALALASERPRO	909
QY	475	GAAGAACCCATTTACGCTCAAGTTGCTTAAAGTAATGCAAAAATTTGACGACTCAAT	534
DB	910	GLUGILUPROILEYTRALAGINVALALALYSEYVALASNLALYSELEAPARGLEUBEN	929
QY	535	CAAAATGCAAGTGGTGGTGTGTTAGGCAAGCAAGCGGCTTCCCTTGAAAGGCGAT	594
DB	930	GINLLEALASERGILYSGILYVALGILGINLALVALGILYPHEPROLEULYSARGHS	949
QY	595	GAT-----	597
DB	950	ASPLYEVALASPALEUSERYVALGILYKRSERVALSERPROGILUPROILEYTRALA	969
QY	598	-----AAAGTTAGATGCTC	612
DB	970	THRLAEAPLEUGILYGLYPROPHEROLEUTHRGHISELYSELYVALASPAERLEU	989
QY	613	AGTAAGTAGAGGCTTCAAGAAATCAAGAAATGCTCAAAAATTTGACATCTCAATCA	672
DB	990	SELYEVALGILYSEUSERYSAEPHEINLYLEUINALGINLYSELEAPHEAENLEUANGIN	1009
QY	673	GCGGTATCGAA 684	
DB	1010	ALVALASERGIN 1013	

RESULT 14
 Q6RRYO_HELPY
 ID Q6RRYO_HELPY PRELIMINARY; PRT; 394 AA.
 AC Q6RRYO;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Cytotoxin associated protein CagA (Fragment).

GN Name=CagA;
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 NC NCBI_TaxID=210;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=ATCC51407;
 RA Yamocka Y., Gutierrez O., Saitou N., Kodama T., Kim J.G., Kashima K.,
 RL Ramirez F.C., Mahachai V., Opat M.S., Graham D.V.,
 DR Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AB057003; BAB87370.1; --; Genomic_DNA.
 DR InterPro: IPR005169; CagA.
 DR Pfam: PF03507; CagA; 1.
 FT NON_TER
 FT NON_TER 394
 SQ SEQUENCE 394 AA; 43035 MW; BIC9D5596ABEB6BF CRC64;

Alignment Scores:
 Pred. No.: 3,86e-62 Length: 394
 Score: 982.50 Matches: 211
 Percent Similarity: 76.9% Conservative: 5
 Best Local Similarity: 75.1% Mismatches: 10
 Query Match: 83.9% Indels: 55
 DB: 2 Gaps: 3

US-09-360-685C-26 (1-685) x Q6RRYO_HELPY (1-394)

QY	1	AAAAATGCAAAATTAAGATTTCAGCAAGGTAAACGACAAAGCAAAAGCACTTGAAAT	60
DB	69	LYSANGILYSAENLYSAEPHESERYVALTHRGINLALYSERAPLEUGLUBEN	88
QY	61	TCGGTTAAAGATGATCATCAATCAAAAGTAACGATTAAGTTGATTCATCAATCA	120
DB	89	SERVALYSAEPVALLEILEANGILYSELETHRAPLYEVALASPAENLEUANGIN	108
QY	121	GCGGTATCAGTGGCTAAAGCAAGGCTGATTCAGTGGGTAGACGACGCTTAGCCGAT	180
DB	109	GLUVALLEUVALALALYSALETHRGILYSAEPHESERMEVALGLUGINLALAEUALAEP	128
QY	181	CTCAAAAATTTCTCAAGAGCAATTGGCCCAACAGCTCAAAAATTAAGATCTCAAT	240
DB	129	LEUASAPHEPHERYSGILUGINLEUINALGINLALYSAENGLUBAPHEAEN	148
QY	241	GCTAGAAAAATCTGAAATATATCAATCCGTTAAGATGCTGATTAAGCACTTAGTC	300
DB	149	ALGILYLYLSESGILUILEYRGINSERVALYSAENGLYVALAENGLYTHRLAEVAL	168
QY	301	GCTAATGGTTATCTCAAGCAAGCAACCACTTTCTTAAAACTTTTGCAATCAAG	360
DB	169	GLYSAENGLYSESGILNALGINLATHRLHLEUSERAPHEAENLEUSERAPILEY	188
QY	361	AAAGATGTGATGCAAACTTGAAATTTCAATTAACATTAACATTAATGACTCAAAAAC	420
DB	189	LYSGILUEUANGILYSEUSERYVALASNLALYSEYVALASNLALYSELEAPARG	206
QY	421	-----GAA	423
DB	207	GLUPROILEYTRALALYSVALASNLYSLEYVALGILGINLALALASERPROGILU	226
QY	424	CCCATTTATGCTAAAGTTAATTAATAAAGAAAGCAAGGCAAGGCTTAGCAAGACCC	483
DB	227	PROILEYTRALALYSVALASNLYSLEYVALGILGINLALALASERPROGILU	246
QY	484	ATTAGGCTCAAGTGGCTTAAAGTAATGCAAAAATTTGACGACTCAATCAATAGCA	543
DB	247	ILEYTRALAGINVALALALYSEYVALASNLALYSELEAPARGLEUBENGINLLEA	266
QY	544	AGTGGTTGGGTGTGTTAGGCAAGCGGCTTCCCTTGAAAGGCAATTAAGTT	603
DB	267	SESGILYSEUGILYVALGILGINLASERGLYPHEPROLEULYSARGHSAPLYEVAL	286
QY	604	GATGATCTCACT-----	615

```

Db      287 AspAsnLeuSerLysValGlyLeuSerAlaArgSerGluProIleTyrAlaThrIleAsp 306
Qy      616 -----AAAGTA 621
Db      307 GlyLeuGlyGlyProIleProLeuLysArgHisAspLysValAlaGluLeuIleLysVal 326
Qy      622 GGGCTTTCAGAGATCAAGAAATGGCTCAGAAATTCAGATCAATCAAGCGGTATCA 681
Db      327 GlyLeuSerArgThrGlnLysLeuAlaGlnLysIleAspAsnLeuAengInAlaValSer 346
Qy      682 GAA 684
Db      347 Glu 347

RESULT 15
Q8KZB0_HELPY
ID      Q8KZB0_HELPY PRELIMINARY; PRT; 327 AA.
AC      Q8KZB0;
DT      01-OCT-2002 (TReMBLrel. 22, Created)
DT      01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT      01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE      Caga (Fragment).
GN      Name=caga;
OS      Helicobacter pylori (Campylobacter pylori).
OC      Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC      Helicobacteraceae; Helicobacter.
OX      NCBI_TaxID=210;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=OK181;
RX      MEDLINE=22295025; Pubmed=12391297; DOI=10.1073/pnas.222375399;
RA      Higashi H., Tsubsumi R., Fujita A., Yamazaki S., Asaka M., Azuma T.,
RA      Hatakeyama M.,
RT      "Biological activity of the Helicobacter pylori virulence factor Caga
RT      is determined by variation in the tyrosine phosphorylation sites.";
RL      Proc. Natl. Acad. Sci. U.S.A. 99:14428-14433(2002).
RN      [2]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=OK181;
RA      Yamakawa A., Yamazaki S., Azuma T.;
RL      Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AB090151; BAC10497.1; -; Genomic_DNA.
DR      InterPro; IPR005169; Caga.
DR      Pfam; PF03507; Caga; 1.
FT      NON_TER 1
FT      NON_TER 327
SQ      SEQUENCE 327 AA; 35798 MW; 1309C5E2418C7DEC CRC64;

Alignment Scores:
Pred. No.: 2.03e-61 Length: 327
Score: 972.50 Matches: 197
Percent Similarity: 90.8% Conservative: 10
Best Local Similarity: 86.4% Mismatches: 20
Query Match: 83.0% Indels: 1
DB: 2 Gaps: 1

US-09-360-685C-26 (1-685) x Q8KZB0_HELPY (1-327)
Qy      1 AAAAATGGCAAAATATAGATTTCAGCAAGGTAAAGCAAAAGCGACCTTGAAT 60
Db      59 LysAengLysAengLysAengLysAengLysValThrGlnAlaLysSerAspLeuGln 78
Qy      61 TCCGTTAAAGATGTGATCATCAATCAAAAGTAAACGATTAAGTTATATCTCAATCAA 120
Db      79 SerIleLysAspValIleIleAengInGlnIleThrAspLysValAspAsnLeuAengIn 98
Qy      121 GGGGTATCAAGTCTAAAGCAAGCGGTGATTTCAAGGTAAAGCAAGCGTTAGCCGAT 180
Db      99 AlaValSerIleAlaAsnSerThrGlyAspPheSerGlyValGlnGlnAlaLeuAlaAsp 118
Qy      181 CTCAAAAATTTCTCAAGAGCAATGGCCCAACAAGCTCAAAAAATGAAAGTCTCAAT 240

```

```

Db      119 LeuLysAsnPheSerLysGlyGlnLeuThrGlnGlnAlaGlnLysAengLysAengLysAeng 138
Qy      241 GCTAGAAAAAATCTGAATATATTCATTCGTTAAAGATGCTGATGAGAACCTTAGTC 300
Db      139 ThrGlyLysAsnSerGluLeuTyrGlnSerValLysAengLysValAengLysThrLeuVal 158
Qy      301 GGTAAATGGGTATCTCAAGCAAGAACCAACCTTCTTAAAAACCTTTCCGACATCAAG 360
Db      159 GlyAsnGlyLeuSerLysAlaGlnAlaThrThrLeuSerLysAsnPheSerAspLys 178
Qy      361 AAAGAGTGAATGCAAAACCTGGAAATTCATTAACATTAACATTAATGACTCAAAAAC 420
Db      179 LysGluLeuAengAlaLysLeuGlnLysAengPheAsnAsnAengAengLysAeng 198
Qy      421 GAACCAATTTATCTTAAAGTTAATTAAGAAAGCAAGCGCAACAGCTAGCCTTGAAGA 480
Db      199 GluProIleTyrAlaLysValAengLysGlnLysThrGlyGlnAlaAengLysGln 218
Qy      481 CCCATTTAGCGCTCAAGTTGCTAAAGGTAAATGCAAAATTCAGCCGACTCAATCAATA 540
Db      219 ProIleTyrThrGlnValAlaLysValLysAlaLysIleAspArgLeuAengLys 238
Qy      541 GCAAGTGTGTTGGGTGTGTAGGGCAAGCAGCGGCTTCCTTTGAAAAGCATGATTA 600
Db      239 AlaSerGlyLeuGlyGlyValGlyGln---AlaGlyPheSerLeuLysGlyHisThrLys 257
Qy      601 GTTGATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAAATGGCTCAGAAATTTG 660
Db      258 ValGlyAspLeuSerLysValGlyLeuSerArgGlnGlnLysGlnLysGlnLysIleAsp 277
Qy      661 AATCTCAATCAAGCGGTATCAAGAA 684
Db      278 AsnLeuAengInAlaValSerGlu 285

```

Search completed: March 6, 2006, 20:00:12
Job time : 223 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 6, 2006, 19:53:26 ; Search time 7.6 Seconds
(without alignments)
1734.434 Million cell updates/sec

Title: US-09-360-685C-26
Perfect score: 1171
Sequence: 1 aaaaatggcaaaaataagga.....caatcaagcgcgtatcagaag 685

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODER=frame+np.model -DEV=xlp
-Q=/abs/ABSSWEB.spool/US09360685/rnat.06032006.121318.15398/app.query.fasta_1
-DB=PIR -QFMT=fasta -SUFFIX=pr -MINMATCH=0.1 -LOOPEXT=0
-UNIT9=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODER=LOCAL
-OUTFMT=pcr -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs802p
-USER=US09360685.QCEN_1_1.63.0rnat.06032006.121318.15398 -NCPU=6 -ICPU=3
-NO_MMAP -NEG SCORES=0 -WAIT -DSBLOK=100 -LONGLOG -DEV TIMEOUT=120
-WAPR TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: PIR.80.*
2: PIR1.*
3: PIR2.*
4: PIR3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1083	92.5	1215	2 B48281	cytochrome-oxidase
2	959.5	81.9	1186	2 C64588	cag pathogenesis
3	896.5	76.6	1167	2 B71924	cag island protein
4	140	12.0	6713	2 B89921	hypothetical prote
5	127.5	10.9	774	2 JC2299	cell surface glyco
6	125	10.7	2481	2 D90011	fmrb protein (limp
7	124	10.6	504	2 D71615	hypothetical prote
8	124	10.6	1302	1 JC6009	surface-located me
9	121	10.3	993	2 C90072	hypothetical prote
10	121	10.3	1125	2 E90598	membrane nucleas
11	121	10.3	2401	2 T28676	thoptery protein -
12	119.5	10.2	821	2 S67087	hypothetical prote
13	119	10.2	615	2 AB2641	methyl-accepting c
14	119	10.2	622	2 D97423	mclA protein (limp

15	118	10.1	1713	2 A55347	adhesive ligand ep
16	117.5	10.0	482	2 C86322	hypothetical prote
17	117	10.0	672	2 S61463	p83/100 protein -
18	117	10.0	693	2 S61464	p83/100 protein -
19	117	10.0	693	2 I40090	p93 protein - Lyme
20	116	9.9	667	2 B97012	methyl-accepting c
21	116	9.9	1175	2 D35815	myosin heavy chain
22	116	9.9	1175	2 C35815	myosin heavy chain
23	116	9.9	1201	2 A35815	myosin heavy chain
24	116	9.9	1201	2 B35815	myosin heavy chain
25	116	9.9	2385	2 A32491	myosin heavy chain
26	116	9.9	2411	2 B32491	myosin heavy chain
27	115.5	9.9	1269	2 B47730	myosin heavy chain
28	115	9.8	1191	2 B97116	probable myosin he
29	114.5	9.8	481	2 T18465	chromosome segrega
30	114.5	9.8	1051	2 T18351	hypothetical prote
31	114.5	9.8	1120	2 A10561	hypothetical prote
32	114.5	9.8	1365	2 T30822	integral membrane
33	114.5	9.8	1939	1 A46762	limp1 protein - Myc
34	114	9.7	1005	2 A46465	myosin alpha heavy
35	114	9.7	3672	2 T23433	hypothetical prote
36	114	9.7	3704	2 T37316	probable laminin a
37	114	9.7	3712	2 S18253	laminin alpha-1 ch
38	113.5	9.7	722	2 H97217	uncharacterized co
39	113.5	9.7	955	1 A35254	leukotoxin A - Pas
40	113.5	9.7	2139	2 T18296	myosin heavy chain
41	113	9.6	1039	2 S62509	probable vesicular
42	113	9.6	1044	2 T50213	mechanosensitive c
43	113	9.6	1120	2 P06693	mechanosensitive c
44	113	9.6	1120	2 B85544	prophage p13 prote
45	113	9.6	1640	2 D86798	

ALIGNMENTS

RESULT 1

B48281

Cytochrome-oxidase associated gene A protein cagA - Helicobacter pylori (strains G39G and CCUG N1) alternate names: Immunodominant 128k antigen; Immunodominant 135k antigen

C1Species: Helicobacter pylori

C1Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C1Accession: B48281; A48281

A1Accession: B48281

A1Molecule type: DNA

A1Residues: 1-1215 <COV1>

A1Cross-References: UNIPROT:P80200; UNIPARC:UPI000017A916; GB:X70039; NID:G394912; PIDN

A1Experimental source: strain G39G

A1Note: sequence extracted from NCBI backbone (NCBIN:133971, NCBI:P.133973)

A1Accession: A48281

A1Molecule type: DNA

A1Residues: 1-957,1026-1215 <COV2>

A1Cross-References: UNIPARC:UPI000002P39F; GB:X70039; NID:G394912; PIDN:CAA49633.1; PI

A1Experimental source: strain CCUG 17874

A1Note: sequence extracted from NCBI backbone (NCBIN:133974, NCBI:P.133968)

C1Comment: These sequences are from cytochrome oxidase producing strains. It is similar to a se

F:958-1025/Region: 33-residue repeats

Alignment Scores:

Pred. No.: 1.9e-70

Score: 1083.00

Percent Similarity: 77.0%

Best Local Similarity: 77.0%

Query Match: 92.5%

DB: 2

Gaps: 1

US-09-360-685C-26 (1-685) x B48281 (1-1215)

QY 1 AAAAATGGCAAAAATTAAGATTTCAGCAGAGTAACGACGCAAAAGCGACTTGTAAT 60

D
750 LysAenGIyLysAnlysaPheSerIyValTrrGlnAlaySeRaspLeuGluAn 769

QY 61 TCGGTAAAGATCTGATCATCAATCAAAGATAACGGATTAAAGTGATATCTCAATCAA 120

Db SerValLySaPaValIlelleSngInLvSvalTrnAsplySvalAspPenLeuAngIn 789

QY 121 GCGGATTCAGTGGCTTAAGACAACGGGTGATTTCACTAGGAGTAGAGACCGTTAGCCGAT 180

Db AlaVaISerValAlayelAlathGlYasPheSerArgValGIudInlAleuAlaAp 809

QY 181 CTCAAAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCAAAAAAAGAAGTCAT 240

Db LeuLySaPePheserIySeluInleuAlagInGlnAlagInleuAbngInUsertLeuAn 829

QY 241 CCTAGAAAAAATCTGAATATATCAATCCGTAAAGATGGTGAATGAAACCGATGC 300

Db AlAyArGylaySeSerIdunIeTyRGInserValLySaMndIYalSaNGIynrLeuAl 849

QY 301 GGTAATGGGTATCTCAGACGAAGCAACTCTTTCTTAAAACTTTTGGCACATCAG 360

Db GlYsaMndIYleuSeRgInlaGlnAlaThrThreusertIyaSpheSerApIlleYs 869

QY 361 AAAGAGTTGATATGCAAAACTTGGAATTTCAATTAACATPAACATATGACTCAAAAAC 420

Db LysGIuLeuSaMnlalyeLyuGIYaSpheSaMnaSaMnaSaMnaSaMnGlyLeuLySaSn 889

QY 421 GAACCCATTTAAGCTTAAAGTTAATTAATAAAGAACGCCGACGACGATGACCTTGAAAGA 480

Db GIuProIIeTyAlalyVaIenlySyLySaMndIYgInAlaAlaIsertLeuGIudIn 909

QY 481 CCCATTTAGCTCAAGTTGGTAAAGTTAAAGTAAATGCAAAAATTTGACGACTCATCAATA 540

Db ProIIeTyAlagInValAlalySyLySaMnaAlayVilLeaPaRgLeuMndInIle 929

QY 541 GCAGAGGTGGTGGTGTTCGACGACGACGCGCGCTTCCTTTGAAAAGCATGATTA 600

Db AlaseRcIyLeuGIyValIyAlGIyInAlaIaGIyPheProLeuLySaRghIAspLy 949

QY 601 GTTGAATGATCTCAATTAAGTAAAGG----- 624

Db ValAspAspLeuSerIySvalGIyAlYargSerValSerProGIuProIIeTyAlaThr 969

QY 624 ----- 624

Db 970 IleAspAspLeuGIyGLyProPheProLeuLySaRghIAspLySvalAspAspLeuSer 989

QY 624 ----- 624

Db 990 LySValGIyArgSerValSerProGIuProIIeTyAlaThrIleAspAspLeuGIyGLy 1009

QY 625 ----- 624

Db 1010 ProPheProLeuLySaRghIAspLySvalAspAspLeuSerIyValLeuSezArgSaIn 1029

QY 637 CAAGATTGGCTCAGAAAATTTGATCAATCTCAATCAAGCGGTATCGCA 684

Db 1030 GInGInLeuAlaGInIyIleAspAspenLeuAngInIlaValaserId 1045

RESULT 2

C64588

cag pathogenicity island protein cag26 - Helicobacter pylori (strain 26695)

N|Alternate names: cytoCoxin-associated gene A protein

I|Species: Helicobacter pylori

C|date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #ext_change 09-Jul-2004

C|Accession: C64588

R|omb, J.F.; White, O.; Kariyaev, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.

Peterson, S.; Lettun, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.; McKenna,

son, J.D.; Kelley, J.M.; Cotton, M.D.; Wetland, J.W.; Fujii, C.; Bowman, C.; Watney, L.

Nature 388, 539-547, 1997

A|Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpi, P.D.; Smith, H.O.; Fraser, C.

A|Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

#Reference number: A65520; MUID:9734467; PMID:9252185

A:Accession: C64588			
A:Status: preliminary, nucleic acid sequence not shown; translation not shown			
A:Molecule type: DNA			
A:Residue 1-1186 x NM			
A:Cross-references: UNIPROT:P55980; UNIPARC:UPI0000126DB8; GB:AE000569; GB:AE000511; NID			

Alignment Scores:			
Pred. No.:	1,41e-61	Length:	1186
Score:	959.50	Matches:	205
Percent Similarity:	78.4%	Conservative:	2
Best Local Similarity:	77.7%	Mismatches:	20
Query Match:	81.9%	Indels:	37
DB:	2	Gaps:	3

US-09-360-685C-26 (1-685) x C64588 (1-1186)			
QY	1	AAAAATGGCAAAAATTAAGATTTTCAGACAGGTAATCGACGCAAAAAGCAGCTTGAAAT	60
DB	755	LYSSENILYLSAASNLVSAAPPHSESLYVALTHRGILHLVSESRPLEUGLUSN	774
QY	61	TCGGTAAAGATGTGATCATCATCAAAAGTAACGGAATTAAGTGAATTCATCA	120
DB	775	SERVALYSAASVALIILEIENGINLVATHRAPHYVALAPHAENLEUENGIN	794
QY	121	CGCGTATCAAGTGGCTAAAGCAACGCGTATTCAGTAGGTAAGCAAGCGTTAGCCAT	180
DB	795	ALALASERVALALALYSALAMETGLYAPPHSESRHVALGILUGINVALLEUHLAASP	814
QY	181	CTCAAAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCAAAAATGAAAGTCTCAT	240
DB	815	LEULYASAPHSESLYSGILUGINLEUHLGILGILNLYSAENGILUSAPHASN	834
QY	241	GCTAGAAAAAATGTGAATATATATCATCCGCTTAAGATGCTGTAATGAAACCTTATC	300
DB	835	THRGILYLSAASNSRGLUENLYRGLINSESLVALYSAASERVALASNLVTHRIEVAL	854
QY	301	GGTAATGGGTTATATTCAGACAGCAACCACTTTCTTAAAACTTTTCGACATCAAG	360
DB	855	GLYASNGLYLEUSERTGLYILEGILALERTHALALUHLALYSAASPHSESRPILLEY	874
QY	361	AAAGGTTGAATGCAAACTTGAAATTTCAATTAACATTAACATTAATGACTCAAAAC	420
DB	875	LYSGILUENLENGILYSPHETLYSAASPHSESRHASNENSNASNNGILYLEULYASN	893
QY	421	-----GAACCCATTTTAACTTAAAGTTAATTAAGAAAGCAAGGCAAGCAAGCTTAC	474
DB	894	SERTHNGULPOLLERYALALYVALAASNLVSLYSTRIDLYGILVALALASERPRO	913
QY	475	GAAGAACCCATTTTAACTTCAAGTTGTAAGTAATTAATGCAAAATTTGACCACTCAT	534
DB	914	GILUGULPOLLERYTHRGILNVALALALYSLYVALAASNLVSLALASERPLEUEN	933
QY	535	CAATTAACAGTGGTTGGTGGTTGTAAGGCAAGCAGCGGCTTCCTTTGAAAGGCAAT	594
DB	934	GILHLEALASERGLYENGILYVALGILGILNLAALGLYPHEPROLEULYASNGHS	953
QY	595	GATTAAGTTGATGATCTCAAGTAGGAGGCTT-----	627
DB	954	ASPHYVALASAPHSESLYVALGILEUSERLALASERPROGLUPOLLERYALA	973
QY	627	-----	627
DB	974	THRILEASAPHSESLYGLYPROPHEROLEULYASNGHSAPHYVALASAPHSELEU	993
QY	628	-----TCAGAAATCAAAATTTGGCTCAAAAATTTGCAATTCATCA	672
DB	994	SERLYVALALYATGSESRHGAENGINGILUENHLGILYSLIEPHASNLUENGIN	1013
QY	673	CGCGTATCAAGA 684	
DB	1014	ALALVALSERGLU 1017	

B71924
 cag island protein, cytotoxicity associated immunodominant antigen - Helicobacter pylori
 C:Species: Helicobacter pylori
 A:Variety: strain J99
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
 C:Accession: B71924
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Moberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 A:Reference number: A71800; MUID:59120557; PMID:9923662
 A:Accession: B71924
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1167 <ARN>
 A:Cross-references: UNIPROT:Q9ZLT1; UNIPARC:UPI0000126BD7; GB:AE001483; GB:AE001439; NID
 C:Experimental source: strain J99
 C:Genetics:
 A:Gene: caga

Alignment Scores:
 Pred. No.: 4,72e-57 Length: 1167
 Score: 896.50 Matches: 189
 Percent Similarity: 74.8% Conservative: 7
 Best Local Similarity: 72.1% Mismatches: 15
 Query Match: 76.6% Indels: 51
 Gaps: 2

US-09-360-685C-26 (1-685) x B71924 (1-1167)

```

QY 1 AAAAATGGCAAAATATAGATTTCAGCAAGTAACGCAAGCAAAAGCGACCTTGAAT 60
DB 754 LysAsnGlyLysAsnLysAsnLysAsnLysValThrGlnAlaLysSerAspLeuGlnAsn 773
QY 61 TCCGTTAAAGATGTATCATCAATCAAAAGTAACGATTAAGTTGATATCATCA 120
DB 774 SerIleLysAspValIleIleLeuGlnLysIleThrAspLysValAsnLeuGln 793
QY 121 GGGGTATCAGTGTACCAACGCGGTATTCAGTGGGTAGACGACGTTAGCCGAT 180
DB 794 AlaValSerValAlaLysAlaThrGlyAspPheSerGlyValGlnGlnAlaLeuAlaAsp 813
QY 181 CTCAAAAATTTCTCAAGAGCAATGGCCCAACAGCTCCAAAAATTAAGTGCAT 240
DB 814 LeuLysAsnPheSerLysGlnLeuAlaGlnAlaGlnLysAsnGlnAsnPheAsn 833
QY 241 GGTAGAAAAATCTGAATATATCAATCCGTTAGAAATGCTGAATCCCTAGTC 300
DB 834 ThrGlyLysAsnSerAlaLeuLysGlnSerValLysAsnGlyValAsnGlyThrLeuVal 853
QY 301 GGTATGGGTATCTCAACGACGACCACTCTTTCAAAAATTTTGGACATCAAG 360
DB 854 GlysAsnGlyLeuSerLysAlaGlnAlaThrThrLeuSerLysAsnPheSerAspIleLys 873
QY 361 AAAGAGTTGATGCAAACTGGAAATTTCAATTAACAATATGACTCAAAAAC 420
DB 874 LysGlnLeuAsnAlaLysLeuGlnLysAsnPheAsnAsnAsnAsnAsnLysLeuGlnAsn 893
QY 421 GAACCATTTATGCTAAATTAATAAAAGAAAGCAAGGCAAGCAAGCTAGCTTGAAGA 480
DB 894 Ser-----ThrGln 896
QY 481 CCCATTAGCTCAAGTTGCTTAAAAAGTAAATGCAAAAATGACCGACTCAATCAATA 540
DB 897 ProIleLysThrGlnValAlaLysLysValLysAlaLysIleAspArgLeuAspGlnIle 916
QY 541 GCAAGTGGTTGGGTGTTGGAGGCAAGCAAGCGGCG----- 576
DB 917 AlaSerGlyLeuGlnLysAspValGlyGlnAlaAlaSerPheLeuLeuLysArgHisAspLys 936
QY 576 ----- 576
DB 937 ValAspAspLeuSerLysValGlyLeuSerAlaAsnHisGlnProIleLysThrAlaThrIle 956

```

QY 577 -----TTCCCTTGAAGAAAGCATGATAAGTTGATGATCTCAGTAG 618
 DB 957 AspAspLeuGlnGlyProPheProLeuLysArgHisAspLysValAspAspLeuSerLys 976

QY 619 GTAGGCGCTTCAAGAAATCAAGATTTGGCTGAGAAATTTGACATCTCAATCAACGGGTA 678
 DB 977 ValGlyLeuSerArgGlnGlnLysLeuThrGlnLysIleAspAsnLeuAsnGlnAlaVal 996

QY 679 TCAGAA 684
 DB 997 SerGln 998

RESULT 4
 B89921
 Hypothesized protein ebha [imported] - Staphylococcus aureus (strain N315)
 C:Species: Staphylococcus aureus
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
 C:Accession: B89921
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogma, A.; Mizutani, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramoto, K. Lancet 357, 1225-1240, 2001
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A:Reference number: A89758; MUID:213111952; PMID:11418146
 A:Accession: B89921
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-6713 <KOR>
 A:Cross-references: UNIPROT:Q99U54; UNIPARC:UPI000011021A; GB:BA000018; PID:g13701232;
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: ebha

Alignment Scores:
 Pred. No.: 0.00934 Length: 6713
 Score: 140.00 Matches: 66
 Percent Similarity: 38.8% Conservative: 50
 Best Local Similarity: 22.1% Mismatches: 97
 Query Match: 12.0% Indels: 86
 Gaps: 2

US-09-360-685C-26 (1-685) x B89921 (1-6713)

```

QY 13 AATAAGATTTCAGCAAGTAACGCAAGCAAAAGCGAC----- 51
DB 5825 AsnLysAspValAspLysGlnValGlnAlaLeuLysAspGluLysAspArgAsnProAsn 5844
QY 52 CTGAAAATTCGTTAAAGATGTATC-----ATCAATCAAAAGTAACGATAAA 102
DB 5845 LeuThrAspLysGlnLysGlnAlaLeuLysAspArgLysGlnGlnIleLeuGlnGlnLys 5864
QY 103 GTTGATATCTCAATCAAGCGGTATTCAGTGTAAAGCAAGCGGTATTCAGTAGGTA 162
DB 5865 HisAsnAspLysAsnAlaLeuThrLysGlnGlnIleGlnAlaLysAlaGlnLeu 5884
QY 163 GAGCAAGCGGTAGCGATCTCAAAAATTTCTCAAG----- 198
DB 5885 AlaGlnAlaLeuGlnAspLysLysAspLeuValLysAlaLysGlnAspAlaLysGlnAsp 5904
QY 199 -----GAGCAATGGCCCAACAGCTCAAAAATTAAGTCAATGCT 243
DB 5905 ValAspLysGlnValGlnAlaLeuLysAspGluLysAspGlnAsnProAsnLeuThrAsp 5924
QY 244 AGAAAAA-----TCGAAATATATCAATCCGTTAAAGATGCT 282
DB 5925 LysGlnLysGlnAlaLeuLysAspArgLysAsnGlnLysGlnGlnLysGlnGlnLys 5944
QY 283 GTGATGAAACCTTAGTCGTTAATGTTATCTCAACGACGACGACCAACTTTCTAAA 342
DB 5945 IleAsnAsnAlaMetThrLysGlnGlnLysGlnAlaLysAla-----GlnLeuAlaGln 5963
QY 343 AACTTTTCGACATCAAG----- 360

```

[illegible]

```

RESULT 5
JC2299
cell surface glycoprotein MSG99 - Pneumocystis carinii
C/Species: Pneumocystis carinii
C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
R/Wada, M.; Nakamura, Y.
DNA Res. 1, 163-168, 1994
A:/title: MSG gene cluster
A:/reference number: JC2299; MIDB:96051989; PMID:8535973
A:/molecule type: DNA
A:/residues: 1-774 <Mad>
A:/cross-references: UNIPROT:Q01817; UNIPARC:UPI00000694BC; GB:DJ1909; GB:DJ7441; NID:g555
G/genes: MSG99
G/superfamily: Pneumocystis carinii major surface glycoprotein MSG100
C/keywords: glycoprotein

Alignment Scores:
Pred. No.:          0.0822
Score:              127.50
Percent Similarity: 41.12%
Best Local Similarity: 22.4%
Query Match:        10.9%
DB:                  2

US-09-360-685C-26 (1-685) x JC2299 (1-774)

Oy      7  GGGAAAATTAGACATTTTCAGCAAGTAAACGCAACAAAAAGCATCTTGAAATTCGTT 66
Db      ||| ::::||| | | | | | | | | | | | | | | | | | | | | | | | |
           GlyLeuAspValValTySerIysValIleuLecIcySAsnAlaMetIleGIuLecIcyS 105
Oy      67 AAAGATGTATCATCAATCCAAAGGTAAACGGATTAAGTTGATATCTGCACAAACGCGTA 126
Db      ||| | | | | | | | | | | | | | | | | | | | | | | | |
           LysArgP-----IleIysAlaIysIleGIuLecIcySAsnIleAspGIuLecIcySLeuIle 123
Oy      127 TCAAGTGGCTAAAGCAACGCGTGAATTC----- 153
Db      ::::| | | | | | | | | | | | | | | | | | | | | | | |
           LeuIleLeuAlaThrIysSerIlePhetyrGIuGIucYsAsnIlytyrPheAspThrCyS 143
Oy      154 -----ACTNAGGGTAGACGACGCTTAGCCGATCTCCAAAATTTCTCAAAGGAG 201
Db      -----ActNAGGGTAGACGACGCTTAGCCGATCTCCAAAATTTCTCAAAGGAG 201

```

Db 14 MecPhepHeuIeRothIySaengInIvalIleapPheCyallePhepHeaRgsp 1
144
QY 202 CAATGG-----GCCCAACAGCTCTCAAAAAATGA-----AGTCAAT 24
Db 164 ThrCyTyrLeuHleIySgInIySgluMeTaRgaEngIValPheLauAgaIleuGly 18
QY 241 GCTGAC---AAAAATCTGAATATATCATCATCGTTAGAAATGCTGAAATGGAACCTTA 29
Db 184 AlaapIleHthrIySserIySaPhegInIySglInIuIleCySproHthrLeu 20
QY 298 GTGCGTATAGGGTATATCTCAACAGACCAACACTCTTTCMAAACTTTTGCAACATC 35
Db 204 IleIySglYIySaPhePheulIaSerLeuCySleuIduIenIleSerPhegluSaPhe 22
QY 358 AAGAAACAGTGAATATGCAAACTTGAAATTTCAATACAAATACATATATGACTCAAA 41
Db 224 AenAenApsIleAparPheCySglYSerIeusenRleuIleApsenIdIyAaGlyS 24
QY 418 AACCAACCCATTATGCTAAAGTAAATATAAAAAAGAAAGGCAAGGCAAGCACTGCTTGA 47
Db 244 GluIduIleCyTyrGluIyLeuGly----- 25
QY 478 GAACCATTTAGCTCAAGTGTCTAAAGAGTAAATACCAAAATATGACATCAATCA 53
Db 253 -----IleTyrSerProLeuIySgluHtySgluIyIySserIyLeuIySAs 27
QY 538 ATAGCAAGTGGTGGGTGGTGTGA-----GGGCAAGCAAGGGGCTTC-----CCTTGA 58
Db 271 LeuCySgInIySglYIleIleTyrSerProIleIySglYpHeaProlIeGlu 29
QY 589 AGGCATGATTAAGTTGAATCATCGATAGAGTGGAGCTT-----TCA 63
Db 291 Aty-----GluueRThrLeuIeugluHtyIleGlyIeugluAenLeuIyTyrIySgluIy 30
QY 631 AGGATATCAAGATTTG 64
Db 309 ArgAenAtrgIyIeu 313

[illegible]

A;Gene: Imp3

A;Genetic code: SGC3

C/Superfamily: surface-located membrane protein Imp3: tetratriconeptide repeat domain

Keywords: duplication; membrane protein
E1-24/Domain: signal sequence

F;25-1302/Product: Surface-located meth...

F;957-992/Domain: tetraatricopeptide repeat homology
surface-located membrane protein Lmp3 #status predicted <MAT>

P:993-1026/Domain: tetra tricopeptide repeat homology <TTP1>

E;1089-1120/Domain: tetratricopeptide repeat homology <IT3>

11-1134-1190/Bomai

Alignment Scores:
Pred. No.: 0.143

Score: 124.00

Percent Similarity: 39.5%

```

best local similarity: 21.0%
Query Match: 10.5%

```

DB: 10.6%

1

US-09-360-685C-26 (1-685) x JC

[illegible]

1 AAAAATGGCAAAATA

```

Db      111:::      :::
356  IyBSeTAantLeuAava

```

03-05-360-685C-26 (1-685) x JC6009 (1-1302)

356 lysSerSerLeuLeuaspMetnclIleSerValagIuGInSerLeuSerLysaspLygclu 375
 61 TCCGTTAAACATGTG-----ATCATCAATCAAAAAGTAACGATTAAGTGAATATTC 114
 376 SerMetGInSerAlaaspLeuLeuaspMetnThrLysLeuIleGluTyrLysGluIleLeu 395
 115 AATCAACCGGTATCATGAGCTTAAGCAACGCGGTGATTCAGTAGGGTGAACACGCTTA 174
 336 AspLysPheaspMetnclIuLysGluAla-----LysPheaspMetncluleGluGlnThrArg 413
 175 GCCGATCTCAAAATTTTCTCAAGACGCAATTG----- 207
 414 LysaspMetncluleuaspMetnThrLysaspGluValLysaspaspProaspThrAlaThrLeu 433
 208 -----GCCCAACAGCGCTCAAAAAGTCAAAAGTCAATGACGTACAAAA 249
 434 ValLysaspLeuThrAspMetnAlaLysaspAlaLysSerValThr--AspSerSerasp 452
 250 AAATCTGAATATAT----- 264
 453 LysSerAspLleIleAlaAlaaspGluAlaLeuIleGluAlaLeuAlaaspAlaaspLys 472
 265 -----CAATCGCTTAAGAGATGTCGATGGAACCTTA 297
 473 AlaLysaspGluValaaspGluAlaaspLysSerLleLysGluGlnLeuaspMetnAlaLeuIle 492
 298 GTCGATTAATGGGATTATCTCAAGCAGACCCCAACTCTTTCTAAAAAATTTCGAC-- 354
 493 -----AepYalaspMetnThrLeuLeuProGlnLeuaspMetnaspMetn 505
 355 -----ATCAAGAAAGAAGTGAATGCAAACTTGGAAATTTCAAT-- 353
 506 AspSerGluIleValLysAlaLysGlnSerLeuaspMetnAlaGluIleThrAspMetnAlaaspLys 525
 394 -----ACAATATACATATATGA----- 411
 526 AlaValaaspMetnaspMetnAlaSerMetGlnSerAlaLysSerSerLeuaspMetnaspLys 545
 412 -----CTCAAAACCAACCCATTATGCTAAAGTTAAATTAAGAAAGACAGGCGCA 462
 546 ValThrLysLysLleGlnaspMetn-----LeuThrGluPheaspMetnaspLysAlaLys 563
 463 GCAGTAGCCTTGAAAG----- 480
 564 PheLysGlnLeuaspMetnThrArgLysaspLleaspMetnPheLeuThrAspaspValLys 583
 481 -----CCATTATAGCTCAATGCTAAAAAGTA--ATCAAAAATATGACCGACATC 531
 584 AspaspMetnProaspMetnThrLysLeuValLysaspLeuThrAspMetnAlaLysaspaspLysValLys 603

```

Oy      532  AATCAAAATGCAAGTGGTTGGGTGTTGAGGCAGACGCGGCTCCCTTGAAGAAG 591
      ::: |||
Db      604  SerValThrIysSerSerAsnIysSerGlnIleIleIleAlaIleAsnAspGluLeuIysGln 623
      ::::|
Oy      592  -----CATGATAAAGTGTATGATCTCATGTAAGTACGGGCTT 627
      |||
Db      624  AAlaLeuAspIysAlaIalysAlaIalysAspGlnIleAspIuAlaIleAsnIys----- 640
      |||
Oy      628  TCAGAGATCAAGATATGGCTGCAGAAAATATGACAAATCTCAACATCCGAGTATCAAA 684
      |||
Db      641  SerIleIysGlnGlnIleuSerAspSerIleThrAsnAlaIleAsnIleuLeuAsnIys 659
      ::::|
RESULT 9
C90072
Hypochemical protein SA2436 (imported) - Staphylococcus aureus (strain N315)
C|Species: Staphylococcus aureus
C|Date: 10-May-2001 #sequence_revision 10-May-2001 #ext_change 09-Jul-2004
C|Accession: C90072
R|Kuroda, M.; Ohta, T.; Uchiyama, I.; Bbba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogu-
ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kato, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A|Title: The whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A|Reference numbers: A69759, WUID:21311952, PMID:11418146
A|Accession: C90072
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-993 <KUR>
A|Cross-references: UNIPROT:Q99QZ5; UNIPARC:UPI00000CABA7; CB:BA000018; PTD:G13702601, P
A|Genetic:
A|Gene: SA2436

```

Alignment Scores:

Pred. No.:
Score:

Percent Similarity

Best Local Similarity

Query Match:

DB:

US-09-360-685C-25

97-7589-000-0000

QY 34 ACG

三

DB 457 Thr

03-09-360-685C-26 (1-685) x C90072 (1-993)

[illegible]

```

Db      585 AsnSerSerLeuValSerIleSerGlnArgIleLeuAspGluLeuAsnLysGlyGlnThr 604
QY      409 GGAAGCTCAAAAGCAAGCCATTATGCTAAAGTTAAATAA-----AGAAAGCA 456
Db      605 AlaLeuSerSer-----ValGlnSerLysLeuSerThrIleAspGlnValIleAsnSer 622
QY      457 GGGGACAGCAGCTAGCCTTGAAGAACCCATTACGCTCAAGTGTCTAAAGTAAAGTAAATGA 516
Db      623 GlyGlnIle-----IleLeuLysAsnGlyLysThr 632
QY      517 AAAATTACCGAGCTCAATCAATACGAAGTGCTTTG-----GGTGT 558
Db      633 ArgIleAspArgLeuGlnThrValLeuProSerIleGlnGlnIleThrIleSerAlaVal 652
QY      559 GTAGGGGACAGCAGCGGGCTTCCCTTTGAAAAGGATGATTAAGTTGATGATTCAGTAAG 618
Db      653 LysAsnAlaGlnAlaAsnProLysValLysSerAspValAlaLysAlaAsnPhe 672
QY      619 GTAGGGCTTTCAGAGATCA-----GAATGGCTCAGAAATTCGACATTCATCA 672
Db      673 Val-----ArgAsnAspLeuProGlnLeuGlnIleArgLeuThrAsnAlaThrAla 689
QY      673 GCGGTATCAGAA 684
Db      690 SerValAsnLys 693

RESULT 10
E90598
membrane nuclease, lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: E90598
R:Champan, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallison, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: E90598
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 11125 <KUR>
A:Cross-references: UNIPROT:Q98PM9; UNIPARC:UPI00000D464B; GB:AL445566; PID:g14090108; E
A:Experimental source: strain UAB CTIP
A:Gene: MYPV 6930
A:Genetic code: SGC3

Alignment Scores:
Pred. No.: 0.236 Length: 1125
Score: 121.00 Matches: 66
Percent Similarity: 39.1% Conservative: 41
Best Local Similarity: 24.1% Mismatches: 107
Query Match: 10.3% Indels: 60
DB: 2 Gaps: 13

US-09-360-685C-26 (1-685) x E90598 (1-1125)
QY      4 AATGCGAAATATAGAT-----TTACGACAGGTAACGACAGCAAAAGC 48
Db      581 AAGGlyGluAsnAspSerLysGlnIleAsnThrSerAsnSerArgIleThrLysAsn 600
QY      49 GACCTTGAAATATCCGTTAAAGATGATCATCAATCAAAAGTAAACGATTAAGTTGAT 108
Db      601 AspLeuArgSerIleGlnLysGlnAsnLeuThrThrLysAsnProSerSerAsnSer 620
QY      109 AATTC-----AATCAAGCGGTATCACTGCTAAAGCAAGC 144
Db      621 AsnValGluThrLysAsnGluThrGlnAsnAsnGluAsnSerSerThrLysLysAspGlu 640
QY      145 GGTGATTTCAATAGGATAGCA-----GGCTTACCGCATCTCAAAAATTTCTCAAG 198
Db      641 IleAspThrSerLysThrGlnAspSerThrAsnSerAsnLeuLysAsnGlnLys 660
QY      199 GAGCAATTGGCCCAACAGCTCAAAAATGAAGTCTCAATCTTGA----- 246

```

```

Db      661 ThrAsnGlnValGluThrLysThrAsnThrGluSerAsnSerSerThrAsnLys 680
QY      247 -----AAAAATCTGAATATATCAATCCGTTAAGATGCTG 285
Db      681 GlnGluGluAsnSerSerThrLysLysGlnIleSerLysSerGluSerSerVal 699
QY      286 AATGGAACCTTACGCTATGCGTTATCTCAAGCAAGAACCAACCTCTTTCTAAAC 345
Db      700 AsnAsnSer-----AsnSerThrAsnLysGlnGlu-----GluAsn 711
QY      346 TTTTCGACATCAAGAAAGTTGAATGCAAACTTGAAATTTCAATAACAATCAAT 405
Db      712 IleAspAsnLysGlnLysGlnIleSerLysSerGluSerAsnValAsnAsnSer 731
QY      406 AATGACATCAAAAGCAAGCCATTATGCTAAAGTTAAATGAAG----- 450
Db      732 ThrAsnThrGlnAsnGlnIleThr---ProGluThrAsnGlnSerGlnAsnValIle 750
QY      451 -----AAAGCAGGGCAGCAGCCTTACCTTGAAGAACCATTTACGCTCAAGTGTAA 504
Db      751 IleGlyLysAsnProAsnAsnGlnSerLeuAsnGlnAsnAlaIleAspValSerAlaLys 770
QY      505 AAGTAAATGCAAAATTAATGACCGACTCAATCAATAGCAAGTGCTTGGCTGTAGGG 564
Db      771 LysVal-----LysIleGlyTyrTrpAsn-----IleAsnGluSerValGlySerSer 787
QY      565 CAGCAGCGGGCTTCCCTTGAAGAG-----CATGATTAAGTTGATGATCTCACT 615
Db      788 AlaSerLysAlaPheLeuValAlaLysValIleAspPheAsnLysLeuAspLeuValGly 807
QY      616 AAGTGGGCTTTCAGAGATCAAGAAATGCTCTCAAAATTCAGATCTCAAT----- 669
Db      808 IleGlyGlyLeuValIleGlnGluThrLeuThrLysIleValGlnGluMetAsnLysLeu 827
QY      670 -----CAGCGGTATCAGAA 684
Db      828 SerLysAspSerSerAspLysTrpValGlnValIleSerGln 841

```

```

RESULT 11
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28676; A45521
R:Simha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A:Title: Comparison of two members of a multigene family coding for high-molecular mas.
A:Reference number: Z20507; MUID:97077455; PMID:8920022
A:Accession: T28676
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: UNIPROT:Q26216; UNIPARC:UPI000017B647; EMBL:U36927; NID:g1041784;
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: A45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KE>
A:Cross-references: UNIPARC:UPI000017B648; GB:M34281

Alignment Scores:
Pred. No.: 0.228 Length: 2401
Score: 121.00 Matches: 50
Percent Similarity: 37.9% Conservative: 47
Best Local Similarity: 19.5% Mismatches: 87
Query Match: 10.3% Indels: 72
DB: 2 Gaps: 9

US-09-360-685C-26 (1-685) x T28676 (1-2401)

```


erge, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, B. W.

A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AB2641

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1615 <KUR>

A:Cross-references: UNIPROT:Q8UHY2; UNIPARC:UPI00001644B8; GB:AE008688; PIDN:AL41544.1;

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: mcIA

A:Map position: circular chromosome

C:Superfamily: Methyl-accepting chemotaxis protein mcPA

Alignment Scores:

Pred. No.:	0.339	Length:	615
Score:	119.00	Matches:	52
Percent Similarity:	42.0%	Conservative:	45
Best Local Similarity:	22.5%	Mismatches:	94
Query Match:	10.2%	Indels:	40
DB:	2	Gaps:	9

US-09-360-685C-26 (1-685) x AB2641 (1-615)

QY 88 AAGTAACGATTAAGTTGATTAATCTCAATCAAGC-----GTATCAGTG 132
DB 226 LysLeuThrAspAlaValLysSerLeuSerAspGlyAspLeuGluThrProIleProTyr 245
QY 133 GCTAAAGCAACGGGTATTTCACTAGGCTAGACCAAGCCTTACCGATCTCAAAATTTT 192
DB 246 AlaThrAsnThrAsnGluPheGlyAsnIleAlaArgAlaLeuValIlePheArgGluAsn 265
QY 193 TCMAAGCAATTTGCC-----CAACAGCTCAAAA 225
DB 266 AlaIleGluLysLeuAlaIleGluGlyLysSerAlaGluLysSerAlaIleGluSer 285
QY 226 AATGAAAGTCTCAATCTAGAAAATCTGAATATATCAATCCCTTAAAGATGCTGTG 285
DB 286 GluArgHisArgAsnAspAlaGluLysGluLysLeuAspGlyGluIleGluPheAlaVal 305
QY 286 AAT-----GGAACCTTAGTGGTAATGGGTTATCTCAAGCAGAGCC 327
DB 306 GlyLuiIleAlaSerGlyLeuGlyArgLysSerArgGlyAspLeuSerArgThrIleGlu 325
QY 328 ACAACTCTTTCTAAAACTTTTCGACATCAAGAAAGAGTGAATGCAAACTTGAAT 387
DB 326 ThrProPheAlaGlyArgLeuAspArgLeuArgThrAspPheAsnGluSerLeuLeuAsn 345
QY 388 TTCATATACAT-----AACATATGAGACTC 414
DB 346 LeuArgAspAlaLeuGlyGluIleArgGluArgThrLeuIleIleGluAsnSerGlyIle 365
QY 415 AAAAGCAACCACTTATGCTAAAGTTAATAAAGAAAGAGGCGACAGCTAGCTT 474
DB 366 GluIleGluLysSerValAspLeuSerLysArgThrGluAsnIleAlaIleSerLeu 385
QY 475 GAAGAACCACTTATGCT-----CAAGTTGCTAAAGTAATGCAAAATATGACCA 528
DB 386 GluGluThrAlaAlaValaGluGluIleThrAlaThrValArgSerSerAlaGluArg 405
QY 529 CTCATCAATATGCAAGTGGTGGTGTGTAAGGCGACAGC-----GGCTTC 579
DB 406 AlaArgGluAlaAsnGluAlaValaArgValThrLysGlnSerAlaAspSerSerGlySer 425
QY 580 CTTTGAAGAAAGCATATTAAGTTGATCTCAGTAAGTA-----GGCTTCAAGAAAT 636
DB 426 ValIleSer-----AsnAlaValaPheAlaMetSerArgIleGluGlyLysArgLys 443
QY 637 ---CAAGATGGCTCAGAAAATTGACATCTC 666

DB 444 IleGluGluIleIleGluValIleAspIle 454

RESULT 14
D97423
mcIA protein [imported] - *Agrobacterium tumefaciens* (strain C58, Cereon)

C:Species: *Agrobacterium tumefaciens*

C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 05-Oct-2004

C:Accession: D97423

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldma

A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium t*

A:Reference number: A97359; MUID:21608551; PMID:11743194

A:Accession: D97423

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1622 <KUR>

A:Cross-references: UNIPROT:Q8UHY2; UNIPARC:UPI00000D183A; GB:AE007869; PIDN:AKK6341.

C:Genetics:

A:Gene: AGR_C_929

A:Map position: circular chromosome

C:Superfamily: Methyl-accepting chemotaxis protein mcPA

Pred. No.:	0.339	Length:	622
Score:	119.00 <td>Matches:</td> <td>52</td>	Matches:	52
Percent Similarity:	42.0% <td>Conservative:</td> <td>45</td>	Conservative:	45
Best Local Similarity:	22.5% <td>Mismatches:</td> <td>94</td>	Mismatches:	94
Query Match:	10.2% <td>Indels:</td> <td>40</td>	Indels:	40
DB:	2	Gaps:	9

US-09-360-685C-26 (1-685) x D97423 (1-622)

QY 88 AAGTAACGATTAAGTTGATTAATCTCAATCAAGC-----GTATCAGTG 132
DB 223 LysLeuThrAspAlaValLysSerLeuSerAspGlyAspLeuGluThrProIleProTyr 252
QY 133 GCTAAAGCAACGGGTATTTCACTAGGCTAGACCAAGCCTTACCGATCTCAAAATTTT 192
DB 253 AlaThrAsnThrAsnGluPheGlyAsnIleAlaArgAlaLeuValIlePheArgGluAsn 272
QY 193 TCMAAGCAATTTGCC-----CAACAGCTCAAAA 225
DB 273 AlaIleGluLysLeuAlaIleGluGlyLysSerAlaGluLysSerAlaIleGluSer 292
QY 226 AATGAAAGTCTCAATCTAGAAAATCTGAATATATCAATCCCTTAAAGATGCTGTG 285
DB 293 GluArgHisArgAsnAspAlaGluLysGluLysLeuAspGlyGluIleGluPheAlaVal 312
QY 286 AAT-----GGAACCTTAGTGGTAATGGGTTATCTCAAGCAGAGCC 327
DB 313 GlyLuiIleAlaSerGlyLeuGlyArgLeuSerArgGlyAspLeuSerArgThrIleGlu 332
QY 328 ACAACTCTTTCTAAAACTTTTCGACATCAAGAAAGAGTGAATGCAAACTTGAAT 387
DB 333 ThrProPheAlaGlyArgLeuAspArgLeuArgThrAspPheAsnGluSerLeuLeuAsn 352
QY 388 TTCATATACAT-----AACATATGAGACTC 414
DB 386 GluIleGluLysSerValAspLeuSerLysArgThrGluAsnIleAlaIleSerLeu 385
QY 475 GAAGAACCACTTATGCT-----CAAGTTGCTAAAGTAATGCAAAATATGACCA 528
DB 415 AAAAGCAACCACTTATGCTAAAGTTAATAAAGAAAGAGGCGACAGCTAGCTT 474
QY 373 GluIleGluLysSerValAspLeuSerLysArgThrGluAsnGluAlaIleSerLeu 392
DB 475 GAAGAACCACTTATGCT-----CAAGTTGCTAAAGTAATGCAAAATATGACCA 528
QY 393 GluGluThrAlaAlaValaGluGluIleThrAlaThrValArgSerSerAlaGluArg 412
DB 529 CTCATCAATATGCAAGTGGTGGTGTGTAAGGCGACAGC-----GGCTTC 579
QY 413 AlaArgGluAlaAsnGluAlaValaArgValThrLysGlnSerAlaAspSerSerGlySer 432

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: March 6, 2006, 20:17:32 ; Search time 2.8 Seconds
(without alignments)
978.592 Million cell updates/sec

Title: US-09-360-685C-26
Perfect score: 1171
Sequence: 1 aaaaatgcgaataaaga.....caatcaagcgatcagaag 685

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 135346 segs, 20000420 residues

Total number of hits satisfying chosen parameters: 270692

Minimum DB seg length: 0

Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ n2p.model -DEV=klp
-O=/abs/ABSGWEE.spool/US09360685/runat_06032006_121329_15688/app_query.fasta_1
-DB=Published Applications AA New -OPMT=faatan -SUFFIX=rapbn -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-USER=US09360685@CGN_1_1_17@runat_06032006_121329_15688 -NCPU=6 -ICPU=3
-NO MPAP -NEB SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA New:
1: /cgn2_6/pdata/2/pubpaa/US08_NEW_PUB.pdp:*
2: /cgn2_6/pdata/2/pubpaa/US06_NEW_PUB.pdp:*
3: /cgn2_6/pdata/2/pubpaa/US07_NEW_PUB.pdp:*
4: /cgn2_6/pdata/2/pubpaa/PCT_NEW_PUB.pdp:*
5: /cgn2_6/pdata/2/pubpaa/US05_NEW_PUB.pdp:*
6: /cgn2_6/pdata/2/pubpaa/US10_NEW_PUB.pdp:*
7: /cgn2_6/pdata/2/pubpaa/US11_NEW_PUB.pdp:*
8: /cgn2_6/pdata/2/pubpaa/US60_NEW_PUB.pdp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1127	96.2	1147	6 US-10-615-668-5
2	896.5	76.6	1167	7 US-11-052-554A-121
3	127	10.8	1290	6 US-10-485-517-141
4	125	10.7	1448	6 US-10-485-517-212
5	120.5	10.3	5024	6 US-10-793-626-2364
6	118	10.1	1713	6 US-10-766-317-2
7	118	10.1	1724	6 US-10-766-317-6
8	118	10.1	3333	6 US-10-766-317-4
9	117	10.0	693	7 US-11-196-475-68

10	115.5	9.9	1095	6	US-10-793-626-3154	Sequence 3154, App
11	114	9.7	3712	7	US-11-019-711-48	Sequence 48, Appl
12	112.5	9.6	611	6	US-10-793-626-2586	Sequence 2586, Ap
13	112	9.6	3712	7	US-11-019-711-51	Sequence 51, Appl
14	111.5	9.5	700	7	US-11-196-475-66	Sequence 66, Appl
15	110.5	9.4	708	7	US-11-196-475-76	Sequence 76, Appl
16	110.5	9.4	739	7	US-11-087-099-12273	Sequence 12273, A
17	109.5	9.4	2087	7	US-11-075-195-285	Sequence 28, Appl
18	109	9.3	2480	6	US-10-995-561-825	Sequence 825, App
19	109	9.3	3116	6	US-10-995-561-826	Sequence 826, App
20	108	9.2	401	6	US-11-097-749-3	Sequence 3, Appl
21	108	9.2	863	7	US-11-097-749-2	Sequence 74, Appl
22	105.5	9.0	700	7	US-11-196-475-74	Sequence 163, App
23	105.5	9.0	1404	6	US-10-878-556A-169	Sequence 663, App
24	104.5	8.9	1654	6	US-10-995-561-663	Sequence 661, App
25	104.5	8.9	1938	6	US-10-995-561-662	Sequence 662, App
26	104.5	8.9	1938	6	US-10-995-561-662	Sequence 660, App
27	104.5	8.9	1954	6	US-10-995-561-660	Sequence 664, App
28	104.5	8.9	1972	6	US-10-995-561-664	Sequence 666, App
29	104.5	8.9	1972	6	US-10-995-561-666	Sequence 667, App
30	104	8.9	2107	6	US-10-995-561-827	Sequence 827, App
31	103	8.8	440	6	US-10-976-933-4	Sequence 2, Appl
32	102.5	8.8	1279	6	US-10-793-626-3188	Sequence 3188, Ap
33	102.5	8.8	440	6	US-10-976-933-4	Sequence 70, Appl
34	102	8.7	663	7	US-11-196-475-70	Sequence 78, Appl
35	102	8.7	663	7	US-11-196-475-78	Sequence 276, App
36	101.5	8.7	380	7	US-11-144-833-13	Sequence 280, App
37	101	8.6	2036	7	US-11-124-368A-276	Sequence 281, App
38	101	8.6	2036	7	US-11-124-368A-280	Sequence 278, App
39	101	8.6	2036	7	US-11-124-368A-281	Sequence 277, App
40	101	8.6	2044	7	US-11-124-368A-278	Sequence 1050, Ap
41	101	8.6	2144	6	US-11-124-368A-277	Sequence 142, App
42	100	8.5	794	6	US-10-689-742-142	Sequence 134, App
43	98.5	8.4	903	6	US-11-074-176-134	Sequence 245, App
44	98.5	8.4	1189	7	US-10-485-517-249	
45	98	8.4	239	6	US-10-485-517-249	

ALIGNMENTS

RESULT 1
US-10-615-668-5
Sequence 5, Application US/10615668
Publication No. US20050276819A1
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappucci, Rino
TITLE OF INVENTION: Helicobacter Pylori CAI Antigen Polynucleotides
FILE REFERENCE: CHIR0337
CURRENT APPLICATION NUMBER: US/10/615,668
CURRENT FILING DATE: 2003-07-08
PRIOR APPLICATION NUMBER: 08/471,491
PRIOR FILING DATE: 1995-06-06
PRIOR APPLICATION NUMBER: 08/256,848
PRIOR FILING DATE: 1994-10-21
PRIOR APPLICATION NUMBER: 09/410,835
PRIOR FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.2
SEQ ID NO 5
LENGTH: 1147
TYPE: PRT
ORGANISM: Helicobacter pylori
US-10-615-668-5
Alignment Scores:
Pred. No.: 5,27e-76
Score: 1127.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Length: 1147
Matches: 228
Conservative: 0
Mismatch: 0

Tue Mar 7 12:58:25 2006

us-09-360-685c-26.rapbn

Page 2

Query Match: 96.2% Indels: 0
DB: 6 Gaps: 0
US-09-360-685c-26 (1-685) x US-10-615-668-5 (1-1167)

```
QY 1 AAAAATGCGAATAAATGATTTTCAGCAAGGTAAGCAAGCAAAACGACCTTGAAAT 60
DB 750 LysaenGIyLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 769
QY 61 TCCGTTAAAGATGATCATCAATCAAAAGGTAAGGTAAGGTAAGGTAAGGTAAG 120
DB 770 SerValLysaenValLysaenValLysaenValLysaenValLysaenValLysaen 789
QY 121 GCGGTATCAGGCTTAAGCAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 180
DB 790 AlaValSerValAlaValSerValAlaValSerValAlaValSerValAlaValSer 809
QY 181 CTCGTTAAAGATGATCATCAATCAAAAGGTAAGGTAAGGTAAGGTAAGGTAAG 240
DB 810 LeuLysaenPheSerLysaenLysaenLysaenLysaenLysaenLysaenLysaen 829
QY 241 GCGTAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 300
DB 830 AlaValSerValAlaValSerValAlaValSerValAlaValSerValAlaValSer 849
QY 301 GGTAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 360
DB 850 GlyaenGIyLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 869
QY 361 AAAGAGTGAATGCAAACTTGAATTCATTAATCAATTAATCAATTAATCAATTAAT 420
DB 870 LysaenGIyLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 889
QY 421 GAACCAATTTATGCTAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 480
DB 890 GluProLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 909
QY 481 CCCATTTAGCTCAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 540
DB 910 ProLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 929
QY 541 GCAAGGCTTTGAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 600
DB 930 AlaSerGIyLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 949
QY 601 GTTGAATGATTCAGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 660
DB 950 ValAspAspLeuSerLysaenLysaenLysaenLysaenLysaenLysaenLysaen 969
QY 661 AATTCATCAAGGCTCAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 684
DB 970 AsnLeuAsnGlnAlaValSerGlu 977
```

RESULT 2

US-11-052-554A-121
; Sequence 121, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 121
; LENGTH: 1167
; TYPE: PRT

i ORGANISM: Helicobacter pylori J99
US-11-052-554A-121

Alignment Scores:

Pred. No.:	4,496-59	Length:	1167
Score:	896.50	Matches:	189
Percent Similarity:	74.8%	Conservative:	7
Best Local Similarity:	72.1%	Mismatches:	15
Query Match:	76.6%	Indels:	51
DB:	7	Gaps:	2

US-09-360-685c-26 (1-685) x US-11-052-554A-121 (1-1167)

```
QY 1 AAAAATGCGAATAAATGATTTTCAGCAAGGTAAGCAAGCAAAACGACCTTGAAAT 60
DB 754 LysaenGIyLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 773
QY 61 TCCGTTAAAGATGATCATCAATCAAAAGGTAAGGTAAGGTAAGGTAAGGTAAG 120
DB 774 SerValLysaenValLysaenValLysaenValLysaenValLysaenValLysaen 793
QY 121 GCGGTATCAGGCTTAAGCAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 180
DB 794 AlaValSerValAlaValSerValAlaValSerValAlaValSerValAlaValSer 813
QY 181 CTCGTTAAAGATGATCATCAATCAAAAGGTAAGGTAAGGTAAGGTAAGGTAAG 240
DB 814 LeuLysaenPheSerLysaenLysaenLysaenLysaenLysaenLysaenLysaen 833
QY 241 GGTAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 300
DB 834 ThrGIyLysaenSerAlaLeuLysaenLysaenLysaenLysaenLysaenLysaen 853
QY 301 GGTAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 360
DB 854 GlyaenGIyLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 873
QY 361 AAAGAGTGAATGCAAACTTGAATTCATTAATCAATTAATCAATTAATCAATTAAT 420
DB 874 LysaenGIyLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 893
QY 421 GAACCAATTTATGCTAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 480
DB 894 Ser-----ThrGlu 896
QY 481 CCCATTTAGCTCAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 540
DB 897 ProLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 916
QY 541 GCAAGGCTTTGAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 576
DB 917 AlaSerGIyLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 576
QY 576 ----- 576
DB 937 ValAspAspLeuSerLysaenLysaenLysaenLysaenLysaenLysaenLysaen 956
QY 577 -----TTCCTTTGAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 956
DB 957 AspAspLeuGIyLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 976
QY 619 GTAGGCTTTCAAGATCAAGATTTGCTCAAGATTTGCTCAAGATTTGCTCAAGATTT 678
DB 977 ValGIyLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 996
QY 679 TCAGAA 684
DB 997 SerGlu 998
```

RESULT 3

US-10-485-517-141
; Sequence 141, Application US/10485517
; Publication No. US20050256239A1

Tue Mar 7 12:58:25 2006

us-09-360-685c-26.rapbn

Page 4

```

Oy 340 AAAAACCTTTGGACATCAAGAAAGAGTTGAATGCAAAACTTGGAAATTTCAATTAACAT 390
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 453 Lys-----GluGlnAlaSerGlnValAspAlaGlnLeuThrGlnGlyAsnGlnAsn 470
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 400 AACCAATATGAGACTCAAAAGCAACCCATTATGCTAAAGTTAATATTAAGAAAGCAGAG 455
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 471 -----IleGluAsnAlaGlnSerIleAspAspValAsnThrAlaLys---Asp 485
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 460 CAGACGACTAGCCTTGAGAAAGCCCATTTACGCTCAAGTGTCTAAAAAGCTAATGCAAA 510
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 486 AsnAlaIleGlnAlaIleAspProIleGlnAlaSerThrAspValLysThrAsnAlaArg 505
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 520 ATTACGCGAATCAATCAATCAATGCAATGATGTTGGGATGTTGTAAGGCAAGCGGGCTTC 570
Db :: ||| :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Oy 506 AlaGluLeuLeuThrGluMet----- 512
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 580 CCTTTGAAAAGCAATGATATGTTGATGATCTCACTGAAGTAGGGCTTTCAAGAAATCAA 630
Db ::|||::: ::|||::: ::|||::: ::|||::: ::|||::: ::|||::: ::|||:::
Oy 513 -----GlnAsnLysIleThrGluIleLeuAsnAsnAsnGluThrThrAsnGln 528
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 640 GAATGGGCTCAAGAAATTTGACATCATCATCAATCAAGCGGTATCAAGA 684
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 529 GluLysGlyAsnAspIleGlyProValArgAlaAlaIleArgLysGluGln 543

```

```

US-10-793-626-2964
? Sequence 2964, Application US/10793626
? Publication No. US20050255478A1
? GENERAL INFORMATION:
? APPLICANT: KIMMERLY WILLIAM JOHN
? TITLE OF INVENTION: STREPHLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
? FILE REFERENCE: P1348015
? CURRENT APPLICATION NUMBER: US/10/793,626
? PRIOR FILING DATE: 2004-03-04
? PRIOR APPLICATION NUMBER: 60/164,258
? PRIOR FILING DATE: 1999-11-09
? NUMBER OF SEQ. ID NOS: 44/2
? SOFTWARE: PatentIn Ver. 2.1
? SEQ. ID NO 2964
? LENGTH: 5024
? TYPE: PRT
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Description of Artificial Sequence: synthetic
? OTHER INFORMATION: amino acid sequence
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (5024)
? OTHER INFORMATION: variable amino acid
US-10-793-626-2964

```

Alignment Scores:	
Pred. No. :	0 0428
Score:	5024
Percent Similarity:	120 50
Best Local Similarity:	41 95
Query Match:	23 24
DB:	10 34
Gaps:	6
Length:	5024
Matches:	622
Conservative:	50
Mismatches:	84
Indels:	71
Gaps:	13

OY 28 AAGTAAACGAGCAAAAAGCAGCTTGAAAAATTCGGTTAAGATGTATC--ATCAAT 84
Db |||||
74 LysleuthrThralalylstheutlallalemlaleusnthLeulaalepleusen 93
OY 85 CAAAGSTAACGANTTAATTACTCATCAACGGTATCGGTGAAGCAGC 144
Db |||||
94 -----ThProglyselualallelythrllalesthrlalshlrArg 110
OY 145 GGTGATTTCAGGGTAGAGCAAGGCATTCGATCAAATAATTTCCAAAGAGCA 204
Db |||||
Oy 111 ThrpevalThr--AlagluginserterylallemnillalemsrslaleethisThr 129

Oy 205 TTGGCCCAACAAAGCTCAAAAAATGAAGT-----CTCAATGCT 2
 Db 130 LeuArgGlnAenIleSerAepnGlnSerValThrAenGlnSerAenThrLeuVal 18
 Oy 244 AGAAAAA-----TCGAATATATCAATCCGTAAAGTGGTGAAT 28
 Db 150 GluProGlnLysGlnIleAlaAethrLeuAlaLeuAenAenAlaLysGlnIleValAen 16
 Oy 289 -----GGAACCTAGTCGGTATAGGGTTATCTCAA---GCAGAACCCAAACTCT 33
 Db 170 GluGlnGlnAlaThrLeuAenAlaAenSerIleAenGlnLysAlaGlnAlaIleLeuThr 18
 Oy 337 TCTAAACCACTTTGGACATCAAGAAAGAGCTG-----AATGCA----- 37
 Db 190 ThrLysAenAlaLeuAenProLysGlnGlnLeuAaArgAlaLysGlnAenAlaAepGln 20
 Oy 376 -----AAACTGGAAATTTCAATTAACAATTAACAATTAATGAGACTCAAA 41
 Db 210 GlnIleAenThrLeuAenGlnLeuThrAepAlaGlnAaArgAenSerGlnLysGlnLeuVal 22
 Oy 418 AACGAAACCACTTTATGCTTAAAGTTAATTAATAAAAGAAACAGAGGACAGCTAGAGCTTGA 47
 Db 230 AenSer-----SerGlnThrArgThrGlnValAlaSerGlnLeuAlaLys 24
 Oy 478 GAACCACTTATGCGTCAAGTTGCTAAAAAGTAATATCAAAATTTGACCGAGCTCATCA 53
 Db 245 -----AlaLysGlnLeuAenLysValMetGlnGlnLeuAenAen 25
 Oy 538 ATAGCAAGCTGTTGGGTGTTGTAGGGCAAGCAGCGGGCTTC----- 57
 Db 258 LeuIleAenGlnLysAenGlnMetIleAenSerSerLysPheIleAenGlnAepAlaAen 27
 Oy 580 -----CTTTGAAAAAGCATGATAA 60
 Db 278 GlnGlnGlnAlaLysSerAenAlaIleAlaSerAlaGlnValLeuLysAenLysSerGln 29
 Oy 601 GTTATGATATCTCAAGTAGAGGAGCTTTTCAAGAAATCAAGAAATGGCTCAAGAAATTCAC 66
 Db 298 AenProGlnLeuAepLysVal-----ThrIleGlnGlnAlaIleAen 31
 Oy 661 AATCTCAATCAAGCGGATATCA 68
 Db 312 AenIleAenSerAlaIleAen 318

RESULT 6
 US-10-766-317-2
 ; Sequence 2, Application US/10766317
 ; Publication No. US2006009630A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Martinovich, M. Peter
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING SQUAMOUS CELL CARCINOMA
 ; FILE REFERENCE: 33628/US/RT/RMS
 ; CURRENT PUBLICATION NUMBER: US/10/766,317
 ; CURRENT FILING DATE: 2004-01-27
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: PatentIn version 3.2
 SEQ ID NO. 2

```

: LENGTH: 1713
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-766-317-2

Alignment Scores:
Pred. No.: 0.0674
Score: 118.00
Percent Similarity: 39.4%
Best Local Similarity: 21.2%
Query Match: 10.1%
DB: 6
Length: 1713
Matches: 55
Conservative: 47
Mismatch: 107
Indels: 50
Gaps: 9

```

13 AATAGGATTTGACGAGGTACGCGACGAAAAAGGAGCCTTGAATAATTCGCTTAAGAT 72

```

Db      300 AsnAsnValAsnValaThrGlnSerAlaLysGlnLeuAspValLysIleLysAsn 319
Qy      73 GTGATCATCAATCAAAAGGTAAAGGATTAAGTTGATATCATCAATCAAGCGGTATCACTG 132
Db      320 ValIleArgAsnValHisIleLeuLeuLysGlnIleSerGlyThrAspGlyGlnLysAsn 339
Qy      133 GCTAAAGCAACGGGTGATTTCACTAGTACG-----GTAGAGCAAGCGTTAGCCGAT 180
Db      340 AsnValProSerGlyAspPheSerArgGlnTrpAlaGlnAlaGlnArgMetCysArgGln 359
Qy      181 CTC-----AAAAATTTCTCAAGAGCAATTTGCCCCCAACGCTCAAAAATAAGTAAAGT 234
Db      360 LeuArgAsnArgAsnPheGlyLysHisLeuArgGlnAlaGlnAlaAspLysArgGlnSer 379
Qy      235 CTCAAATGCTGAAAAAATCTGAATATATCATATCCGTTAAGATGCTGAATGGAACC 294
Db      380 GlnLeuLeuLeuAsnArgIleArgGlnTrpGlnLysThrHisGlnGlnLysAsn----- 397
Qy      295 CTAGTCGTAATAGGTTATCTCAAGAGCAAGCAACACTCTTTCTAA-----AAC 345
Db      398 -----AsnGlyLeuAlaAsnSerIleArgAspSerLeuAsnGlnLysGlnAlaLys 414
Qy      346 TTTTGGACATCAAGAAAGAGTGAATGCAAAACCTTGAATTTCAATTAACAATCAAT 405
Db      415 LeuSerAspLeuArgAlaArgLeuGlnGlnAlaAlaGlnAlaLysGlnAlaAsnGly 434
Qy      406 AATGACTCAAAAACGACCCACTTTATGCTAAAGTTAATAAAG----- 450
Db      435 LeuAsnGlnGlnAsnGlnArgAlaLeuGlnAlaIleGlnArgGlnLysGlnLysAsn 454
Qy      451 -----AAGCAGGCGACACTAGCTTGAAGAA 480
Db      455 SerLeuGlnSerAspPheThrLysTrpLeuThrThralaAspSerLeuLeuGlnThr 474
Qy      481 CCCATTACGCTCAAGTTGCTAAAGGTAAATGCAAAATGACCGACTCAATCAATA 540
Db      475 AsnIleAlaLeuGlnLeuMetGlnLysSerGlnLysGlnLysLeu----- 491
Qy      541 GCAAGTGTGGGTGTTGTAGGGCAAGCAGCGGGCTTCCCTTGAAGAAAGCATGATAA 600
Db      492 AlaAlaSerLeuAsnGlnLysAlaArgGln-----GlnLeuSerAspLys 505
Qy      601 GTTATGATCTCAGT-----AAGTAGGGCTTTCAAGAAAT----- 636
Db      506 ValArgGlnLeuSerArgSerAlaGlyLysThrSerLeuValGlnGlnAlaGlnLysHis 525
Qy      637 -----CAAGATTTGCTCAGAAAATTTGACATCTCAATCAAGCGGTATCA 681
Db      526 AlaArgSerLeuGlnGlnLeuAlaLysGlnLeuGlnGlnLysArgAsnAlaSer 544

RESULT 7
US-10-766-317-6
; Sequence 6, Application US/10766317
; Publication No. US20060009630A1
; GENERAL INFORMATION:
; APPLICANT: Marinkovich, M. Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING SQUAMOUS CELL CARCINOMA
; FILE REFERENCE: 33828/US/RFT/RMS
; CURRENT APPLICATION NUMBER: US/10/766,317
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 1724
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-766-317-6

Alignment Scores:
Pred. No.: 0.0674 Length: 1724
Score: 118.00 Matches: 55
Percent Similarity: 39.4% Conservative: 47

```

```

Best Local Similarity: 21.2% Mismatches: 107
Query Match: 10.1% Indels: 50
DB: 6 Gaps: 9

US-09-360-685C-26 (1-685) x US-10-766-317-6 (1-1724)

Qy      13 AATAAGATTTAGCAGAGTAAAGCGAAGCAAAAGCGACTTGAATAATCCGTTAAGAT 72
Db      311 AsnAsnValAsnValaThrGlnSerAlaLysGlnLeuAspValLysIleLysAsn 330
Qy      73 GTGATCATCAATCAAAAGGTAAAGGATTAAGTTGATATCATCAATCAAGCGGTATCACTG 132
Db      331 ValIleArgAsnValHisIleLeuLeuLysGlnIleSerGlyThrAspGlyGlnLysAsn 350
Qy      133 GCTAAAGCAACGGGTGATTTCACTAGTACG-----GTAGAGCAAGCGTTAGCCGAT 180
Db      351 AsnValProSerGlyAspPheSerArgGlnTrpAlaGlnAlaGlnArgMetCysArgGln 370
Qy      181 CTC-----AAAAATTTCTCAAGAGCAATTTGCCCCCAACGCTCAAAAATAAGTAAAGT 234
Db      371 LeuArgAsnArgAsnPheGlyLysHisLeuArgGlnAlaGlnAlaAspLysArgGlnSer 390
Qy      235 CTCAATGCTGAAAAAATCTGAATATATCATATCCGTTAAGATGCTGAATGGAACC 294
Db      391 GlnLeuLeuLeuAsnArgIleArgGlnTrpGlnLysThrHisGlnGlnLysAsn----- 408
Qy      295 CTAGTCGTAATAGGTTATCTCAAGAGCAAGCAACACTCTTTCTAA-----AAC 345
Db      409 -----AsnGlyLeuAlaAsnSerIleArgAspSerLeuAsnGlnLysGlnAlaLys 425
Qy      346 TTTTGGACATCAAGAAAGAGTGAATGCAAAACCTTGAATTTCAATTAACAATCAAT 405
Db      426 LeuSerAspLeuArgAlaArgLeuGlnGlnAlaAlaGlnAlaLysGlnAlaAsnGly 445
Qy      406 AATGACTCAAAAACGACCCACTTTATGCTAAAGTTAATAAAG----- 450
Db      446 LeuAsnGlnGlnAsnGlnArgAlaLeuGlnAlaIleGlnArgGlnLysGlnLysAsn 465
Qy      451 -----AAGCAGGCGACACTAGCTTGAAGAA 480
Db      466 SerLeuGlnSerAspPheThrLysTrpLeuThrThralaAspSerLeuLeuGlnThr 485
Qy      481 CCCATTACGCTCAAGTTGCTAAAGGTAAATGCAAAATGACCGACTCAATCAATA 540
Db      486 AsnIleAlaLeuGlnLeuMetGlnLysSerGlnLysGlnLysLeu----- 502
Qy      541 GCAAGTGTGGGTGTTGTAGGGCAAGCAGCGGGCTTCCCTTGAAGAAAGCATGATAA 600
Db      503 AlaAlaSerLeuAsnGlnLysAlaArgGln-----GlnLeuSerAspLys 516
Qy      601 GTTATGATCTCAGT-----AAGTAGGGCTTTCAAGAAAT----- 636
Db      517 ValArgGlnLeuSerArgSerAlaGlyLysThrSerLeuValGlnGlnAlaGlnLysHis 536
Qy      637 -----CAAGATTTGCTCAGAAAATTTGACATCTCAATCAAGCGGTATCA 681
Db      537 AlaArgSerLeuGlnGlnLeuAlaLysGlnLeuGlnGlnLysArgAsnAlaSer 555

RESULT 8
US-10-766-317-4
; Sequence 4, Application US/10766317
; Publication No. US20060009630A1
; GENERAL INFORMATION:
; APPLICANT: Marinkovich, M. Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING SQUAMOUS CELL CARCINOMA
; FILE REFERENCE: 33828/US/RFT/RMS
; CURRENT APPLICATION NUMBER: US/10/766,317
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT

```


; PRIOR FILING DATE: 2001-09-10
 ; PRIOR APPLICATION NUMBER: 60/318,700
 ; PRIOR FILING DATE: 2001-09-12
 ; NUMBER OF SEQ ID NOS: 227
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO: 48
 ; LENGTH: 3712
 ; TYPE: PRF
 ; ORGANISM: Drosophila melanogaster
 US-11-019-711-48

Alignment Scores:
 Pred. No.: 0.13 Length: 3712
 Score: 114.00 Matches: 67
 Percent Similarity: 35.0% Conservative: 46
 Best Local Similarity: 20.7% Mismatches: 100
 Query Match: 9.7% Indels: 110
 Gaps: 15

US-09-360-685C-26 (1-685) x US-11-019-711-48 (1-3712)

```

QY 13 AATAGGATTTTCAGCAAGTAAAGCAAGCAAAAGCGAGCTTGAATAATTCGTTAAAGAT 72
DB 2348 AmsSerLySpheAepThrValSerGluInLysLeuGlnAlaGluLysAenIleLysAep 2367
QY 73 GTG-----ATCATCAAT-----CAAAAGCTACCGATTAAGTTGAT 108
DB 2368 AlaGlyAenPheLeuIleAenGlyAepLeuThrLeuAenGlnIleAenGlnLysLeuAep 2387
QY 109 AATTCATCAAGCGGATCA-----ACCGTACGCGTAAATGAG----- 129
DB 2388 AenLeuAArgAepAlaLeuAenGlnLysAenSerPheAenLysAenValAepGluInLys 2407
QY 130 -----GTGCTAAAGCAAGCGGATTTTCAGTACGAGGTAG 165
DB 2408 ProValAArgLysAepGlnIleLysGlnIleAenPheAenLysAenValAepGluInLys 2425
QY 166 CAAGCGTTACCGGATTTTCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2445
DB 2426 GlnLysAAlaGlnLysAAlaIleLysAAlaGlnLysAAlaIleLysAAlaGlnLysAAla 2465
QY 226 AATGAAAGTCTCAAT---GCTAGAAAATAATCTGAATATATCAATCCGTTAAGAAATGT 282
DB 2446 ThrAlaSerAlaGluProAlaIleLysAAlaIleThrAlaLysSerGlyIleValGlnAla 2465
QY 283 CTGAATGGA-----ACCGTACGCGTAAATGAG----- 309
DB 2466 ValGlnAlaAlaGlnLysLeuSerGlnAepAlaIleSerAlaAlaGlnLysAAlaIleThrA 2485
QY 310 -----TTATCT 315
DB 2486 LysThrAepGlyIleGlnLysAArgAlaIleLysAAlaAepThrGlySerThrAepLeu 2505
QY 316 CAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 315
DB 2506 GlnAArgAlaArgGlnSerLeuGlnLysAAlaIleLysAAlaIleThrAlaLysSerGlyIle 2525
QY 376 -----AACTGGAAATTTCAATACAAATAAATGAGCTGAAA 417
DB 2526 SerAlaGlyLysValGlnLysIleSerAlaValAenAenAlaIleThrGlnIleGlnLys 2545
QY 418 AACGAAACCATTTATGCTTAAGTTAATAA----- 447
DB 2546 Aep-----IleAenLysLeuIleAepGlnLysProAlaGlnLysGln 2559
QY 448 -----AAGAAAGCAAGCGGCAAGCAAGCTTGA----- 477
DB 2560 ArgAepMetTrpLysAenSerAenAlaAenAlaSerAepAlaLeuGlnIleLysLysAen 2579
QY 478 -----GAACGATTTAGCGCTCAAGTTCGTTAAAG----- 507
DB 2580 ValLeuGlnIleLeuGlnIleProValSerValGlnThrProLysGlnLysLeuGlnLysAlaIle 2599
  
```

```

QY 508 ---GTAAATGCAAAATTGACGAGCTCAATCA-----ATAGCAAGTGTGGGTGTTGA 561
DB 2600 GLyIleAenAArgAepLeuAepLeuThrAenLysAepValSerGlnAlaAenLysGlnLeu 2619
QY 562 GCGCAAGCAGCGGCTTCCTTGAATAAGCAATGAATGATGATCTCACT----- 615
DB 2620 AepAepValGlnLysSerValSerLysLeuAenGlnLysAAlaGlnLysPheIleGlnLys 2639
QY 616 -----AAGTACGGCTTTCAAGCAATCAAGAAATTTGCTGCAAAATTTGCAATTCAT 669
DB 2640 GlnIleAArgValGly---SerGlnSerAenGlnLysGlnLysIleGlnLysLeuLys 2658
QY 670 CAAGCGGCT 678
DB 2659 AlaGlnVal 2661

RESULT 12
US-10-793-626-2586
; Sequence 2586, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2586
; LENGTH: 611
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-2586

Alignment Scores:
Pred. No.: 0.176 Length: 611
Score: 112.50 Matches: 63
Percent Similarity: 38.9% Conservative: 37
Best Local Similarity: 24.5% Mismatches: 91
Query Match: 9.6% Indels: 66
Gaps: 12

US-09-360-685C-26 (1-685) x US-10-793-626-2586 (1-611)
QY 40 GCAAAAGCAAGCTTGAATAATTCGTTAAGTGTGATCAATCAAAAGGTAAAGGAT 99
DB 163 AlaLysAepAlaIleLysThrIle---AepValLeuValLysGlnIleLysAep 181
QY 100 AAGTTATATCTCAATCAAGCGGTATCAAGTGGCTAA-----GCAACGGGTGATTC 153
DB 182 ---IleAepSerAenAenGlnLysThrGlnAArgLysAAlaIleLysAAlaIleGlnLys 200
QY 154 ACTAGGCTAGCAAGCGGTACCGCATCAAAAATTTCTCAAG----- 198
DB 201 GlnAArgLysLeuLysGlnAlaIleAepLysValThrIleSerLysSerIleLysAepIle 220
QY 199 -----GAGCAATTTGGCCCAAGCAAGCTCAAAAATAATGAAAGT 234
DB 221 GlnThrValLysAArgThrAepPheGlnLysIleAepGlnIleAepProLysAArgPheThr 240
QY 235 CTCAATCTTGAATAAATA-----TCTGAATATATCAATCCGTTAAGATGCTGTG 285
DB 241 LeuAenLysAAlaLysLysAepIleIleThrAepValAenThrGlnIleGlnAenGlyPhe 260
QY 286 AATGAAACCGTACGTTAAGTGTATTCATCAAGCAAGCAAGCAAGCAAGCAAGCTTTTAAAAAC 345
DB 261 LysGlnIleGlnLysIleLysGlyLysLeuThrAenGlnLysThrGlnPheAepLysGln 280
  
```

```

QY 346 TTTTCGACATCAAGAA-----GAG 366
Db 281 LeuThrAlaLeuGlnGlySGluPheLeuGluValGluHisAlaHisLeuValGlu 300
QY 367 TTGATGCAAACTTGAAATTTCAATTAACAATTAAGACTCAAAAGCAACC 426
Db 301 LeuAsnGlnLeuGlnGlnGlnPheAsnAsnArgTyrGluHisLeuAsnGlnAlaHis 320
QY 427 ATT-----TATGCTAAAGTTAATAAAG 450
Db 321 LeuLeuGlnGluValHisAlaGluHisLeuLeuGlnGlnValValAlaGlnLeuThr 340
QY 451 AAGCA-----GGGCAAGCAGTACCTTGAAGAACCACTTACGCTCAAGTTGCT 501
Db 341 GlnGlnIleLeuAsnAsnGlnSerAlaSerTyrPheIleLeuGlnTrpAla----- 357
QY 502 AAAAAGTAAATGCAAAATTTGACCGACATCAATTAAGCAAGTGTGGTGTGTA 561
Db 358 -----LeuAspArgIleLeuGlnIleGln-----LeuGluTrpMet 369
QY 562 GGGCAAGCAGCGGCTTCCTTTGAAAGCAGATGAATTAAGTTGATGCTCAAGTAAAGTA 621
Db 370 AsnSerIleArgGly-----AlaHisThrValGlnAspValHisLeuVala 384
QY 622 GGGCTTTCAAGATCAAGATTTGGCTCAGAAATTTGACATCAATC 670
Db 385 LeuLeu-GlnGlyIleGlnGlnIleLeu--LysValAsnValSerIle 399

RESULT 13
US-11-019-711-51
; Sequence 51, Application US/11019711
; Publication No. US2006009634A1
GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernyev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Baturajan, Meera
; APPLICANT: Grose, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Sciore, Paul
; APPLICANT: Eilerman, Karen
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Rothenberg, Mark
; APPLICANT: Stone, David J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Anderson, David W
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Miller, Charles E.
; APPLICANT: Eisen, Andrew J
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-235
; CURRENT APPLICATION NUMBER: US/11/019,711
; PRIOR FILING DATE: 2004-12-21
; PRIOR APPLICATION NUMBER: US/10/037,417
; PRIOR FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/260,018
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: 60/260,360
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/272,411
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/272,817
; PRIOR FILING DATE: 2001-03-02

```

```

; PRIOR APPLICATION NUMBER: 60/291,186
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 60/303,231
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/305,060
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/318,405
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/318,700
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 227
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 51
; LENGTH: 3712
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; US-11-019-711-51

Alignment Scores:
Pred. No.: 0.182 Length: 3712
Score: 112.00 Matches: 68
Percent Similarity: 33.7% Conservative: 41
Best Local Similarity: 21.1% Mismatches: 104
Query Match: 9.6% Indels: 110
DB: 7 Gaps: 14

US-09-360-685C-26 (1-685) x US-11-019-711-51 (1-3712)

QY 13 AATAAGATTTCAGACAGTAAGCAAGCAAAAGCCTTGAATAATTCCTTAAGAT 72
Db 2348 AsnSerIlePheAsnThrValSerGlnGlnValGlnLysAsnIleLysAsp 2367
QY 73 GTG-----ATCATCAAT-----CAAAAGTAAAGTAAAGTTGAT 108
Db 2368 AlaGlyAsnPheLeuIleAsnGlnLysPheThrLeuAsnGlnIleAsnGlnLysAsp 2387
QY 109 AATTCATCAAGCGGTATCA----- 129
Db 2388 AsnLeuArgAspAlaLeuAsnGlnLeuAsnSerPheAsnLysAsnValAspGlnGluLeu 2407
QY 130 -----GTGGCTAAAGCAAGCGGTATTCAGTAGGTTAGAG 165
Db 2408 ProValArgGluAspGlnHisLeuGlnLysAlaAspAlaLeuThrAsp-----GlnAlaGln 2425
QY 166 CAAGCGTTAGCGCATCTCAAAATTTCTCAAGGAGCAATTGCCCAAGCTCAAAA 225
Db 2426 GlnLysAlaAlaGlnLeuAlaIleLysAlaGlnAspLeuAlaAlaGlnIleThrAspMet 2445
QY 226 AATGAAGTTCAT--GCTAGAAAAAATCTGAATATATCAATCCGTTAAGATGCT 282
Db 2446 ThrAlaSerAlaGluProAlaIleLysAlaAlaThrAlaTyrSerGlyIleValGlnLys 2465
QY 283 GTGAATGGA-----ACCTAGTCGTTAATGG----- 309
Db 2466 ValGlnAlaAlaGlnLysLeuSerGlnAspAlaIleSerAlaAlaGlnLysAlaThrAsp 2485
QY 310 -----TTACT 315
Db 2486 LysThrAspGlyIleGlnGlnArgAlaHisLeuAlaAspThrGlySerThrAspLeuLeu 2505
QY 316 CAAGCAAGCCACACTTTCTTAAAACTTTTCGACATCAAGAAAGCTTAATGCA 375
Db 2506 GlnArgAlaArgGlnSerLeuGlnLysValGlnAspLeuGlnLysProArgLeuAsnAla 2525
QY 376 -----AACTTGGAATTTCAATTAACAATTAAGCTCAAA 417
Db 2526 SerAlaGlyLysValGlnLysIleSerAlaValAsnAsnAlaThrGlnHisGlnLeuLys 2545
QY 418 AACGAACCATTTATGTAAAGTTAATAA----- 447
Db 2546 Asp-----IleAsnLysLeuIleAspGlnLeuProAlaGlnLysGln 2559
QY 448 -----AAGAAAGCAGGGCAAGCAGCTAGCTTGA----- 477

```

Tue Mar 7 12:58:25 2006

us-09-360-685c-26.rapbn

Page 10

```

Db 2560 ArgAspMetTrpValAsnSerValAsnAlaAsnAlaSerAspAlaLeuGluIleLeuValAsn 2579
Qy 478 -----GAAACCATTTACGCTCAAGTTCCTAAAG----- 507
Db 2580 ValLeuGluIleLeuGluIleProValSerValGluThrProValSerGluLeuGluValAlaHis 2599
Qy 508 ----GTAAATGCAGAAATATGACCACTCAATCAAAATAGCAAGTGTGGGT----- 555
Db 2600 GlyIleAsnArgAspLeuAspLeuThrAsnValAspValSerGluAlaAsnValGluLeu 2619
Qy 556 -----GTTGTAGGCGCAAGACGCGGCTCCCTTTGAAAAGCATGATAAGTTGATGAT 609
Db 2620 AspAspValGluGluSerValSerValSerGluLeuAlaGluAspIleGluGluGln 2639
Qy 610 CTGATGAGTATGAGGCTTTCAAGGAATCAAGATTGGCTCAGAAAATGACATCTCAAT 669
Db 2640 GluHisArgValGly---SerGlnSerArgGlnLeuGluIleGluIleGluAsnLeuVal 2658
Qy 670 CAAGCGGTA 678
Db 2659 AlaGlnVal 2661

RESULT 14
US-11-196-475-66
/ Sequence 66, Application US/11196475
/ Publication No. US20050271682A1
/ GENERAL INFORMATION:
/ APPLICANT: Dattwyler, Raymond J.
/ APPLICANT: Gomes Solecki, Maria J. C.
/ APPLICANT: Luft, Benjamin J.
/ APPLICANT: Dunn, John J.
/ TITLE OF INVENTION: Recombinant Construct of Borrelia
/ FILE REFERENCE: 2631.1001-011
/ CURRENT APPLICATION NUMBER: US/11/196,475
/ PRIOR FILING DATE: 2005-08-03
/ PRIOR APPLICATION NUMBER: US 08/148,191
/ PRIOR FILING DATE: 1993-11-01
/ PRIOR APPLICATION NUMBER: US 08/235,836
/ PRIOR FILING DATE: 1994-04-29
/ PRIOR APPLICATION NUMBER: US 09/666,017
/ PRIOR FILING DATE: 2000-09-19
/ PRIOR APPLICATION NUMBER: US 60/226,484
/ PRIOR FILING DATE: 2000-08-18
/ PRIOR APPLICATION NUMBER: PCT/US01/24736
/ NUMBER OF SEQ ID NOS: 213
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 66
/ LENGTH: 700
/ TYPE: PRT
/ ORGANISM: Borrelia burgdorferi
US-11-196-475-66

Alignment Scores:
Pred. No.: 0,208 Length: 700
Score: 111,50 Matches: 50
Percent Similarity: 42,2% Conservative: 59
Best Local Similarity: 19,4% Mismatches: 68
Query Match: 9,5% Indels: 81
Db: 7 Gaps: 10

US-09-360-685c-26 (1-685) x US-11-196-475-66 (1-700)
Qy 13 AATAAGATTTCAGCAGATGACGCAAGCAAGCAAGCAAGCTTGAATAATTCGTTAAAGAT 72
Db 303 AspValAlaGlnGluIleLeuAspSerAlaGluAspAsnLeuAsp-----ValGlnArg 320
Qy 73 GTGATCATCATCAATCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 132
Db 321 AsnThrValArgGluValIleGlnGluIleAspIleAsnGluIleAsnValGluValAsnLeu 340

Qy 133 GCTAAAGCAACGGGTGATTTCAAT-----AGGTAGAGCA 168
Db 341 ProVal---ProGluAspValSerSerProValAspValSerGlnLeuGlnIleValGlu 359
Qy 169 GCGTATGCCGATTTCTCAAAATTTCTCAAGAGCAATTTGCCCAACAAAGCTCAAAAAAT 228
Db 360 SerLeuGluAspLeu-----GlnGlnGluLeuValSerGluThrGluValArgGluAsn 375
Qy 229 GAAAGTCTCAATGCTAGAAAAAATCTGAATATATATCAATCCGTTAAGATGGGTGAT 288
Db 376 GlnValArgGluIleGluValSerGlnIleGluIleValSerValSerAspGluVal 392
Qy 289 GAAACCTTAGTCGGTATATGCTTATCTCAAGCAGAACCAACTTTCTTAAAAACCTT 348
Db 393 -----LeuLeuValSerValAspAspValValAspSerValSerValSerValSerVal 405
Qy 349 TCG-----GACATCAAGAAAGATGATGCAAAA----- 378
Db 406 ValAlaLeuAspLeuAspArgGluLeuAsnSerValAspSerValSerValSerValSerVal 425
Qy 379 -----CTTGAAATTTCAATAC 396
Db 426 AlaValSerGluGluIleThrValSerGluValSerGlnValSerLeuGluValSerValSerVal 445
Qy 397 AATTAACATTAATGAGCTCAAAAACGA-----CCCATTAATGCTAAATTAATAA 447
Db 446 AspGluAsnLeuMetMetProGluAspGlnValSerProGluValValSerValSerValSer 465
Qy 448 AAGAAAGCAGGCGACAGCTAGGCTTGAAGAACCATTTACGCTCAAGTTGCTAAAG 507
Db 466 ValSerValSerProValSer----- 473
Qy 508 GTAATGCAAAATTCACGACTCAACAAATGCAAGTGGTGGTGTGAGGCA 567
Db 474 -----GluValGluValLeuAspValLe 481
Qy 568 GCAAGCGCTTCCTTTGAAAGGCTATTAAGTTGATATCTCAATGAGTGGCTT 627
Db 482 -----PheValSerValSerValSerValSerValSerValSerValSerValSerVal 496
Qy 628 TCAAGATCAATATGCTCAAGAAATGCAATCTCAATCAAGCGGTATCA 681
Db 497 SerSerValSerValSerValSerValSerValSerValSerValSerValSerValSerVal 514

RESULT 15
US-11-196-475-76
/ Sequence 76, Application US/11196475
/ Publication No. US20050271682A1
/ GENERAL INFORMATION:
/ APPLICANT: Dattwyler, Raymond J.
/ APPLICANT: Gomes Solecki, Maria J. C.
/ APPLICANT: Luft, Benjamin J.
/ APPLICANT: Dunn, John J.
/ TITLE OF INVENTION: Recombinant Construct of Borrelia
/ FILE REFERENCE: 2631.1001-011
/ CURRENT APPLICATION NUMBER: US/11/196,475
/ PRIOR FILING DATE: 2005-08-03
/ PRIOR APPLICATION NUMBER: US 08/148,191
/ PRIOR FILING DATE: 1993-11-01
/ PRIOR APPLICATION NUMBER: US 08/235,836
/ PRIOR FILING DATE: 1994-04-29
/ PRIOR APPLICATION NUMBER: US 09/666,017
/ PRIOR FILING DATE: 2000-09-19
/ PRIOR APPLICATION NUMBER: US 60/226,484
/ PRIOR FILING DATE: 2000-08-18
/ PRIOR APPLICATION NUMBER: PCT/US01/24736
/ NUMBER OF SEQ ID NOS: 213
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 76
/ LENGTH: 708
/ TYPE: PRT
```

ORGANISM: Borrelia burgdorferi
US-11-196-475-76

Alignment Scores:

Pred. No.:	0.246	Length:	708
Score:	110.50	Matches:	51
Percent Similarity:	44.1%	Conservative:	54
Best Local Similarity:	21.4%	Mismatches:	94
Query Match:	9.4%	Indels:	39
DB:	7	Gaps:	9

US-09-360-685C-26 (1-685) x US-11-196-475-76 (1-708)

```
OY 13 AATAAGATTTCAGAGGTAACGCAAGCAAAAAGCCTTGAAATTCGGTTAAGAT 72
DB 303 AspLysAlaGlnGlnLysLeuAspSerAlaGluAspAsnLeuAsp-----ValGlnArg 320
OY 73 GTGATCATCAATCAAAAAGGTAACGATTAAGTTGATATCTCAATCAACGGTATCATGTG 132
DB 321 AspThrValArgGlnLysIleGlnGlnAspIleAsnGlnIleAsnLysGlnLysAsnLeu 340
OY 133 GCTAAGCAACGGGTATTTTCACT-----AGGTTAGACGCA 168
DB 341 ProLys--ProGlyAspValSerProLysValAspLysGlnLeuGlnIleLysGln 359
OY 169 GCGTTAGCCGATCTCAAAAATTTCTCAAGAGCAATTCGCCCAACAGCTCAAAAAT 228
DB 360 SerLeuGlnAspLeu-----GlnGlnGlnLeuLysGlnIleGlyAspGlnAsn 375
OY 229 GAAAGTCTCAATGCTAAGAAAAA-----TCTGAATATATCA 267
DB 376 GlnLysArgGlnIleGlnLysGlnIleGlnIleLysLysArgAspGlnGlnLeuLeuLys 395
OY 268 TCCGTTAAGATGCTGATGATGACACCTAGTCGGTATGCGTTATCTCAAGCAGAAAGCC 327
DB 396 SerLysAspGlnLysValSerLysAspTyrGlnAlaLeuAspLeuAspArgGlnLeuSer 415
OY 328 ACAACTCTTTTAAAAAATTTTCGACATCAAGAAAGAG-----TTG 369
DB 416 LysAlaSerSerLysGlnLysSerLysValLysGlnGlnIleThrLysGlnLysSer 435
OY 370 AATGCAAACTTGAAATTTCAATACATTAACAATATGACTCAAAAACGACCCATT 429
DB 436 ArgAlaSerLeuGlnLysLeuAsnAspLysAsnLeuMetLeuProGlnAsp----- 453
OY 430 TATGCTAAAGTAATTAAGAAAGAGGAGGCAAGCAAGCTAGCTGAAGAACCATTTAC 489
DB 454 -----GlnLysLeuProGlnAspLysLysLeuAspSerLysLeuAsp 467
OY 490 GCTCAAGTTGCTAAAAAGTAATGCAAAAATTTGACCCGACTCAATCAATAGCA--AGT 546
DB 468 GlnLysLysGlnPheLysProValSerGlnValGlnLysLeuAspLysLysSer 487
OY 547 GGTTCGGTGTGTAGGCAAGCAGCGGCTTCCCTTGAAAAAGCATGATTAAGTTGAT 606
DB 488 AsnAsnAsnGlnValGlnLysLeuSerProLeuAspLysProSerTyrAspAspIleAsp 507
OY 607 GATCTCAGTAGAGTGGCTTCAAGATCAAGATCAAGATTCAGAAAAATTGAC 660
DB 508 -----SerLysGlnGlnValAspAsnLysAlaIleAsnLeuGlnLysIleAsp 523
```

Search completed: March 6, 2006, 20:22:43
Job time : 32 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: March 6, 2006, 20:04:12 ; Search time 28.7 Seconds
(without alignments)

1994.516 Million cell updates/sec

Title: US-09-360-685c-26

Perfect score: 1171

Sequence: 1 aaatgcgaataaataaga.....caatcagcgcgcatcagaag 685

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODU=frame.n2p.model -DEV=xlp
-O=/abs/ABSSWEB_spool/US09360685/runat_06032006_121327_15626/app_query.fasta_1
-DB=Published Applications AA Main -QFMT=fastaan -SUPFIX=rapbm -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 -UNITS=bite -START=1 -END=1 -MATRIX=blomsu62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -HOST=abs802p
-USER=US09360685@CGN_1_1_307@runat_06032006_121327_15626 -NCPU=6 -ICPU=3
-NO_WMAP -NEG_SCORES=0 -WAIT -DSPBLOC=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications AA Main:

1: /cgn2_6/ptodata/1/pubpa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpa/US10_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1127	96.2	1147	3	US-09-921-157-5
2	1127	96.2	1338	3	US-09-402-100-4
3	896.5	76.6	1167	4	US-10-335-977-8633
4	896.5	76.6	1178	4	US-10-335-977-8634
5	896.5	76.6	1183	4	US-10-335-977-8635
6	140	12.0	6713	4	US-10-282-122A-43811
7	131.5	11.2	1404	5	US-10-732-923-3304
8	131	11.2	3533	4	US-10-282-122A-70177
9	129	11.0	1992	5	US-10-470-048B-81
10	129	11.0	10498	5	US-10-470-048B-440
11	127	10.8	837	3	US-09-815-242-5883

12	127	10.8	875	3	US-09-815-242-13080	Sequence 13080, A
13	127	10.8	1361	4	US-10-369-493-3209	Sequence 3209, Ap
14	127	10.8	2434	3	US-09-815-242-5835	Sequence 5835, Ap
15	127	10.8	6281	3	US-09-815-242-12996	Sequence 12996, A
16	126	10.8	2368	3	US-09-815-242-5635	Sequence 5635, Ap
17	126	10.8	2368	3	US-09-815-242-12389	Sequence 12389, A
18	125.5	10.7	2437	3	US-09-815-242-5834	Sequence 5834, Ap
19	125	10.7	2478	3	US-09-815-242-5816	Sequence 5816, Ap
20	125	10.7	2478	3	US-09-815-242-12967	Sequence 12967, A
21	125	10.7	2478	5	US-10-470-048B-220	Sequence 220, App
22	125	10.7	2481	4	US-10-282-122A-43762	Sequence 43762, A
23	124	10.6	5795	3	US-09-815-242-12610	Sequence 12610, A
24	123	10.5	10203	4	US-10-661-809-23	Sequence 23, Appl
25	123	10.5	10203	4	US-10-724-972A-4098	Sequence 4098, Ap
26	122.5	10.5	998	4	US-10-282-122A-70540	Sequence 70540, A
27	121.5	10.4	6641	4	US-10-282-122A-70560	Sequence 70560, A
28	121	10.3	993	4	US-10-282-122A-43875	Sequence 43875, A
29	120	10.2	724	4	US-10-282-122A-71401	Sequence 71401, A
30	120	10.2	1090	4	US-10-282-122A-56217	Sequence 56217, A
31	120	10.2	3158	3	US-09-815-242-12611	Sequence 12611, A
32	118	10.1	1441	5	US-10-732-923-3352	Sequence 3352, Ap
33	118	10.1	1693	4	US-10-603-725-4	Sequence 4, Appl1
34	118	10.1	1693	4	US-10-603-725-8	Sequence 8, Appl1
35	118	10.1	1713	4	US-10-171-311-113	Sequence 113, App
36	118	10.1	1713	4	US-10-372-683-10	Sequence 10, Appl
37	118	10.1	1713	4	US-10-603-725-6	Sequence 6, Appl1
38	117.5	10.1	1724	4	US-10-603-725-2	Sequence 2, Appl1
39	117.5	10.0	1047	4	US-10-282-122A-44353	Sequence 44353, A
40	117.5	10.0	2025	3	US-09-815-242-5703	Sequence 5703, Ap
41	117	10.0	693	4	US-10-369-100-68	Sequence 68, Appl
42	116	9.9	2067	6	US-11-097-143-40167	Sequence 40167, A
43	116	9.9	2539	5	US-10-831-070-6	Sequence 6, Appl1
44	115	9.8	873	3	US-09-952-267-13	Sequence 13, Appl
45	115	9.8	873	5	US-10-872-768-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-921-157-5

Sequence 5, Application US/09921157
Publication No. US20040048353A1

GENERAL INFORMATION:

APPLICANT: Covacci, Antonio

APPLICANT: Bugnoli, Massimo

APPLICANT: Telford, John

APPLICANT: Macchia, Giovanni

APPLICANT: Rappelli, Rino

TITLE OF INVENTION: Helicobacter Pylori Cytotoxin Proteins Useful For

FILE REFERENCE: CHIR0315

CURRENT APPLICATION NUMBER: US/09/921,157

PRIOR FILING DATE: 2001-08-02

PRIOR APPLICATION NUMBER: 08/466,662

PRIOR FILING DATE: 1995-06-06

PRIOR APPLICATION NUMBER: 08/256,848

PRIOR FILING DATE: 1994-10-21

PRIOR APPLICATION NUMBER: 09/360,934

PRIOR FILING DATE: 1999-07-26

PRIOR APPLICATION NUMBER: 08/471,491

PRIOR FILING DATE: 1995-06-06

PRIOR APPLICATION NUMBER: PCT/EP93/00472

PRIOR FILING DATE: 1993-03-02

PRIOR APPLICATION NUMBER: PCT/EP93/00158

PRIOR FILING DATE: 1993-01-25

PRIOR APPLICATION NUMBER: FI 92 A 000052

NUMBER OF SEQ ID NOS: 10

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 5

LENGTH: 1147

TYPE: PRT

ORGANISM: Helicobacter pylori

Tue Mar 7 12:58:25 2006

УВ-09-360-685с-26. рабм

Page 2

US-09-921-157-5

Alignment Scores:

Pred. No.:	1,3e-96	Length:	114
Score:	1127.00	Matches:	228
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	96.2%	Indels:	0
DB:		gaps:	0

US-09-360-685C-26 (1-685) x US-09-921-157-5 (1-1147)

Qy	1	AAAAATGCGAAAATAAGATTTCAGCAGAGTAAACGACAGCAAAAGCGACTTGAAAT	60
Db	750	LYSAENGLYVSAENLYSAPHSESTERYETALTHRGINALYSESTASPLEUGLUN	761
Qy	61	TTCGGTTAAAGATGTGATCATCAATCAAAAGTAAACGGATAAGGTATATCTCAATCA	121
Db	770	SERVALYSAPPEVALIETLEAENGLNLYVALTHRAPELYSVALASPEMLEUANGIN	785
Qy	121	GGCGTATCACTGGCTAAAGCAACGGGTGATTCTATAGGGTAGACGAACGGTTAGCCAT	181
Db	790	ALVALSERVALALALYSAETHRGILASPPHESERTAYVALGLNGLNLALEVALASPP	805
Qy	181	CTCGAAATTTCTCAAGAGCAATGGCCCAACAGCTCAAAAAATGAAAGTGCAT	241
Db	810	LEULYAPHPHESERTYSELUGLNULENLAGINGINLAGLINLYSAPENGLINSESTEN	825
Qy	241	GCTAAGAAAAATCTGAATATATCAATCCGTATAGATGGTGTAATGGACCCATATC	301
Db	830	ALALAGLYSAPSESTENILETRYGINSEVALYLSAENGLYVALAENGLYTHRIEVAL	849
Qy	301	GGTAAATGGATTCTCAAGCAGAAACCAATCTCTTCTTAAAACTTTTGGACATCAAG	361
Db	850	GLYAPENGLYSESTENLAGLUNLATHTHRIEUSESTYASPPHESERTAPILLEYLS	865
Qy	361	AAAGAGTTGAATGCCAAACTTGAAATTTCAATATCAATATCAATATATGACCTCAAAAC	421
Db	870	LYSGULNEUNSNALALYSENGLYAPENPHEANSNAPNSAPNSAPNSAPNSAPNSAPNS	885
Qy	421	GAAACCATTTATGCTAAAGTTAATATAAAAAGAAACAGGGCAACAGCTAGCTGTAGAA	481
Db	890	GLNPPROLETRYALALYSAVALAENLYSYLSYSEVALGLYGLINLAALASSTENGLN	905
Qy	481	CCCATTTATGCTCAAGTTGCTAAAAAGTAAATGCAAAAATTGACCGACATCAATAA	541
Db	910	PROLETRYALAGLIVALALALYSLYSAVALAENGLALYSLILEAPARGLEUNSNGLILE	925
Qy	541	GCAAGTGGTTGGTGGTGTATGAGGGCAACAGCGGGCTTCCCTTGAAAGAGGTATGA	601
Db	930	ALASERTLYENGLYVALVALGLYGLINLAALGLYPHEPROLEULYBARGILASPPYLS	945
Qy	601	GTGTATGATCTCAAGTAAAGTAAAGGCTTTCACAGAAATCAAGAAATGGCTAGAAATGAC	661
Db	950	VALASPPHESERLYSVALGLYSESTERYAPENGLINGLNULENLSELINLYSILEAP	965
Qy	661	AATTCATCAACGCGGTATCAGA 684	
Db	970	ASMPENAGLINALALASSTENGLN 977	

RESULT 2

US-09-402-100-4

sequence 4, Application US/09402100
Patent No. US20010010000

NO. 0520010019834A1
; GENERAL INFORMATION:

APPLICANT: Daewoong

APPLICANT: Kim, Byung-O

APPLICANT: YII, Young-Hye

APPLICANT: Park, Myung-Hwa

APPLICANT: Choi, Deok-Joon

APPLICANT: Jung, Hyung-Jin

STATE OF INVENTION: Recomb

```

? FILE REFERENCE: 0136/003140
? CURRENT APPLICATION NUMBER: US/09/402,100
? CURRENT FILING DATE: 1999-09-27
? EARLIER APPLICATION NUMBER: KR 97-11950
? EARLIER FILING DATE: 1997-03-31
? EARLIER APPLICATION NUMBER: KR 97-11951
? NUMBER OF SEQ ID NOS: 48
? SOFTWARE: Patentin version 3.0
? SEQ ID NO 4
? LENGTH: 1338
? TYPE: PRT
? ORGANISM: Artificial Sequence
? FEATURES:
? OTHER INFORMATION: CsgA/CtxA2B chimeric protein
? US-09-402-100-4

```

Alignment Scores:

Pred. No.:	1,32e-96	Length:	1338
Score:	1127.00	Matches:	228
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	96.2%	Indels:	0
DB:	3	Gaps:	0

US-09-360-685C-26 (1-685) X US-09-402-100-4 (1-1338)

[illegible]

Db 970 AsnLeuasnGlnAlaValSerGln 977

RESULT 3
US-10-335-977-8633
Sequence 8633, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandagouras, Amy B.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 8633:
SEQUENCE CHARACTERISTICS:
LENGTH: 1167 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (8) LOCATION 1...1167
SEQUENCE DESCRIPTION: SEQ ID NO: 8633:
US-10-335-977-8633

Alignment Scores:
Pred. No.: 4.65e-75 Length: 1167
Score: 896.50 Matches: 189
Percent Similarity: 74.8% Conservative: 7
Best Local Similarity: 72.1% Mismatches: 15
Query Match: 76.6% Indels: 51
Gaps: 2

US-09-360-685C-26 (1-685) x US-10-335-977-8633 (1-1167)

QY 1 AAAAATGCAAAATTAAGATTTCAGCAAGTAAAGCAAAAGCGACTTGAAT 60
Db 754 LysaAngLyLysaAsnLysaPheSerValThrGlnAlaLysaSerAspLeuGln 773
QY 61 TCCGTTAAAGATGATCAATCAATCAAAAGTAAAGTAAAGTGAATTCATCA 120
Db 774 SerLeLysaPheValIleLeuAsnGlnLysIleThrAspLysValAspAsnLeuAsnGln 793
QY 121 GCGGATACATGCGTAAAGCAAGGATTCAGTAGGAGTAAAGCAAGCGTAC 180
Db 794 AlaValSerValAlaLysAlaThrLysaPheSerGlyValGlnGlnAlaLeuAlaAsp 813

QY 181 CTCAAAAATTTCTCAAGAGCAATGGCCCAACAGCTCAAAAAATGAATCTCAT 240
Db 814 LeuLysaAsnPheSerLysGlnGlnLeuAlaGlnGlnAlaGlnLysaAsnGlnAspPheAsn 833
QY 241 GCTAGAAAAAATCTGAATATATCATCCGTTAGAAATGGTGGCAATGAACCTATC 300
Db 834 ThrGlyLysaSerAlaLeuLysGlnSerValLysaAngLyLysaGlyThrLeuVal 853
QY 301 GGTAAATGGTTATCTCAAGACAGAACCAACTTTCTTAAAACTTTCCGACATCAAG 360
Db 854 GlyaAngLyLeuSerLysAlaGlnAlaThrThrLeuSerLysaAsnPheSerAspLys 873
QY 361 AAAGGTGAATGCAAAACCTTGAAATTCATATCAATCAATCAATATGACTCAAAAAC 420
Db 874 LysGlnLeuAsnAlaLysLeuGlyAsnPheAsnAsnAsnAsnAngLyLeuGlnAsn 893
QY 421 GAACCATTTATGCTAAAGTTAATTAAGAAACAGGCGCAAGCAGTACCTGAAGAA 480
Db 894 Ser-----ThrGln 896
QY 481 CCCATTACGCTCAAGTTGCTTAAAGGTAATGCAAAATTTGACCGACTCATCAATA 540
Db 897 ProLeuLysaThrGlnValAlaLysLysValLysAlaLysIleAspArgLeuAspGlnIle 916
QY 541 GCAAGTGTGGTGGTGTGAGGCGCAAGACGGGC----- 576
Db 917 AlaSerGlyLeuGlyAspValGlyGlnAlaLaserPheLeuLeuLysArgHisAspLys 936
QY 576 ----- 576
Db 937 ValAspAspLeuSerLysValGlyLeuSerAlaAsnHisGlnProLeuLysAlaThrIle 956
QY 577 -----TTCCCTTTGAAAAGCATGATAAGTTGATGATTCAGTAAG 618
Db 957 AspAspLeuGlyGlyProPheProLeuLysArgHisAspLysValAspAspLeuSerLys 976
QY 619 GTAGGCGCTTCAAGGAATCAAGAAATGGCTCAGAAAATTTGACATTCATCAAGCGTA 678
Db 977 ValGlyLeuSerArgLysGlnLysLeuThrGlnLysIleAspAsnLeuAsnGlnAlaVal 996
QY 679 TCAGAA 684
Db 997 SerGln 998

RESULT 4
US-10-335-977-8634
Sequence 8634, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997

Tue Mar 7 12:58:25 2006

us-09-360-685c-26.rapbm

Page 4

```

ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 742-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 8634:
SEQUENCE CHARACTERISTICS:
LENGTH: 1178 amino acids
TYPE: amino acid
MOLECULE TYPE: linear
TOPOLOGY: linear
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...1178
SEQUENCE DESCRIPTION: SEQ ID NO: 8634:
US-10-335-977-8634

Alignment Scores:
Pred. No.: 4,666-75 Length: 1178
Score: 896.50 Matches: 189
Percent Similarity: 74.8% Conservative: 7
Best Local Similarity: 72.1% Mismatches: 15
Query Match: 76.6% Indels: 51
Gaps: 2

US-09-360-685C-26 (1-685) x US-10-335-977-8634 (1-1178)
QY 1 AAAAAAGCAAAATTAAGATTTCAGCAAGGTAACGCAAGCAAAAGCACTTGAAT 60
DB 765 LysaenGlyLysaenLysaenPheSerLysValThrGlnAlaLysSerPheLysLys 784
QY 61 TCCGTTAAAGATGTGATCATCATCAATCAAAAGGTAACGTAAGTGAATTCATCA 120
DB 785 SerLeuLysaenPheValLleLeuGlnLysLysLysLysLysLysLysLysLys 804
QY 121 GCGATATCAGTGTGTAAGCAACGCGGATTCAGAGGTAAGGTAAGCAAGCGTGA 180
DB 805 AlaValSerValAlaLysAlaThrGlyAspPheSerGlyValGlnGlnAlaLeuAla 824
QY 181 CTCAAAATTTCTCAAAAGAGCAATGCGCCAAACAAGCTCAAAAATGAAGTCTCA 240
DB 825 LeuLysaenPheSerLysGlnLysLysLysLysLysLysLysLysLysLysLys 844
QY 241 GCTAGAAAAATCTGAAATATATCAATCCGTTAAGATGAGTGAATGAACCTGAT 300
DB 845 ThrGlyLysaenSerAlaLeuLysGlnSerValLysaenGlyValaLysnLys 360
QY 301 GGTATGGGTTATCTCAGCAAGGCAACCACTCTTTCTTAAAACTTTTCGACATCA 360
DB 865 GlyaenGlyLysSerLysAlaGlnAlaThrLysSerLysaenPheSerAlaLys 884
QY 361 AAAGGTGAAGTGAAGCAATTCGAAATTCATTAACATTAACATTAATGACTCA 420
DB 885 LysGlnLysaenAlaLysLysLysLysLysLysLysLysLysLysLysLysLys 904
QY 421 GAACCATTTATGCTAAAGTTAATTAAGAAAGCAAGGCAAGCAAGCTTGAAGA 480
DB 905 Ser-----Thru 907
QY 481 CCCATTAGCGCTCAAGTTGCTTAAAGGTAATGCAAAATGACCGACTCAATCAATA 540
DB 908 ProLeuLysThrGlnValAlaLysLysValLysLysLysLysLysLysLysLysLys 927
QY 541 GCAAGTGTGGTGGTGGTGGTGGGCAAGCAAGCGGCGC----- 576
DB 928 AlaSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 947
QY 576 ----- 576

ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 742-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 8635:
SEQUENCE CHARACTERISTICS:
LENGTH: 1183 amino acids
TYPE: amino acid
MOLECULE TYPE: linear
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...1183
SEQUENCE DESCRIPTION: SEQ ID NO: 8635:
US-10-335-977-8635

Alignment Scores:
Pred. No.: 4,666-75 Length: 1183
Score: 896.50 Matches: 189
Percent Similarity: 74.8% Conservative: 7
Best Local Similarity: 72.1% Mismatches: 15
Query Match: 76.6% Indels: 51
Gaps: 2

RESULTS
US-10-335-977-8635
Sequence 8635, Application US/10335977
Publication No. US2004005279A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSER: LAHYE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1975
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 742-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 8635:
SEQUENCE CHARACTERISTICS:
LENGTH: 1183 amino acids
TYPE: amino acid
MOLECULE TYPE: linear
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...1183
SEQUENCE DESCRIPTION: SEQ ID NO: 8635:
US-10-335-977-8635

DB 948 ValaAspLysSerLysValGlyLysSerAlaAsnHISGluProLeuThrLys 967
QY 577 -----TTCCCTTGAAGGCAATGAAGTGAATGATTCGTTAG 618
DB 968 AspAspLysGlyGlyProPheProLeuLysArgHisAspLysValaAspLysSer 987
QY 619 GTAGCGCTTTCAGGAATTCAGCAATTCGCAAAATGCAATTCATCATCAAGCGTA 678
DB 988 ValGlyLysSerArgGlnGlnLysLysLysLysLysLysLysLysLysLysLys 1007
QY 679 TCAGAA 684
DB 1008 SerGlu 1009
```

```

US-09-360-685c-26 (1-685) x US-10-335-977-8635 (1-1183)
QY 1 AAAAATGGCAAAATAGCATTTTCAGCAAGGTACGCAAGCAAAAGCCGACTTGAAT 60
Db 770 LysAsnGlyLysAsnLysAspPheSerLysValThrGlnAlaLysSerAspLeuGln 789
QY 61 TCCGTTAAAGATGTGATCATCATCAAAAGGTAAAGGATTAAGTATATATCATCA 120
Db 790 SerLleLysAspValIleIleAsnGlnLysIleThrAspLysValAspAsnLeuAsn 809
QY 121 GCGGATTCAGTGGCTAAAGCAAGCGGTGATTTTCAGTAGGGTAGAGCAAGCGTGA 180
Db 810 AlaValSerValAlaLysAlaThrGlyAspPheSerGlyValGluGlnAlaLeuAla 829
QY 181 CTCAAAATTTCTCAAAAGCAAGCAATGGCCCAACAGCTCAAAAAATGAATCTCAAT 240
Db 830 LeuLysAsnPheSerLysGlnLysIleGlnGlnAlaGlnLysAsnGlnAspPheAsn 849
QY 241 GCTAGAAAAAATCTGAATATATATCAATCGTTAAGATGGTGAATGGAACCCAGTC 300
Db 850 ThrGlyLysAsnSerAlaLeuLysValLysAsnGlyValAsnGlyThrLeuVal 869
QY 301 GGTATGGGTTATCTCAAGCAAGGACCAACTCTTTCTAATAAATTTTCGACATCAAG 360
Db 870 GlyAsnGlyLeuSerLysAlaGlnAlaThrThrLeuSerLysAsnPheSerAspLys 889
QY 361 AAAGGTTGAATGCCAAACTTGAATTTTCATTAACATTAACATTAATGACTCAAAAC 420
Db 890 LysGlnLeuAsnAlaLysLeuGlyAsnPheAsnAsnAsnAsnAsnGlyLeuGlnAsn 909
QY 421 GAACCATTTATGCTAAAGTTATATAAAGAAAGCAAGGCAAGCAGTACCTTGAAGA 480
Db 910 Ser-----ThrGln 912
QY 481 CCCATTAGCTCAAGTCTAATAAAGGTAATGCAAAATTTGACCGACTCATCAATA 540
Db 913 ProIleLysThrGlnValAlaLysValLysValLysIleAspArgLeuAspGlnIle 932
QY 541 GCAAGTGGTTGGGCTTGTGAGGCAAGCAAGCGGCG- 576
Db 933 AlaSerGlyLeuGlyAspValGlyGlnAlaAlaSerPheLeuLeuLysArgHisAspLys 952
QY 576 ----- 576
Db 953 ValAspAspLeuSerLysValGlyLeuSerAlaAsnHisGlnProIleLysAlaThrIle 972
QY 577 -----TTCCCTTTGAAAGGCAATGATTAAGTGTGATGATCTCAGTAG 618
Db 973 AspAspLeuGlyGlyPheProheProLeuLysArgHisAspLysValAspAspLeuSerLys 992
QY 619 GTAGGCTTTCAAGCAATCAAGATTGGCTCAGAAATTTGACAAATCTCAATCAAGCGGTA 678
Db 993 ValGlyLeuSerArgLysGlnLysLeuThrGlnLysIleAspAsnLeuAsnGlnAlaVal 1012
QY 679 TCAGAA 684
Db 1013 SerGln 1014

```

```

RESULT 6
US-10-282-122A-43811
; Sequence 43811, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert

```

```

APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELTRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 43811
LENGTH: 6713
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-10-282-122A-43811

```

```

Alignment Scores:
Pred. No.: 0.00033 Length: 6713
Score: 140.00 Matches: 66
Percent Similarity: 38.8% Conservative: 50
Best Local Similarity: 22.1% Mismatches: 97
Query Match: 12.0% Indels: 86
DB: 4 Gaps: 11
US-09-360-685c-26 (1-685) x US-10-282-122A-43811 (1-6713)
QY 13 AATAAGATTTCAGCAAGGTAAAGCAAAAGCCAGC----- 51
Db 5825 AsnLysAspValAspLysGlnValGlnAlaLeuIleAspGlnLysAspArgAsnProAsn 5844
QY 52 CTGAAATTCGTTAAAGATGTGATC-----ATCATCAAAAGGTAAAGCAATATA 102
Db 5845 LeuThrAspLysGlnLysGlnAlaLeuLysAspArgLleAsnGlnIleLeuGlnGln 5864
QY 103 GTTGAATATCTCATCAAGCGGTATCAAGTGGCTTAAGCAAGCGGTGATTTCAAGAGGTA 162
Db 5865 HisAsnAspIleAsnAsnAlaLeuThrLysGlnLysIleGlnGlnAlaLysAlaGlnLeu 5884
QY 163 GAGCAAGGTTAGCGCATCTCAAAATTTCTCAAG----- 198
Db 5885 AlaGlnAlaLeuGlnAspIleLysAspLeuValLysAlaLysGlnAspAlaLysGlnAsp 5904
QY 199 -----GAGCAATGGCCCAACAGCTCAAAATTAAGTGTCAATGCT 243
Db 5905 ValAspLysGlnValGlnAlaLeuIleAspGlnLysAspGlnAsnProAsnLeuThrAsp 5924
QY 244 AGAAAAA-----TCTGAATATATCATCTCGTTAAGATGCT 282
Db 5925 LysGlnLysGlnAlaLeuLysAspArgLleAsnGlnIleLeuGlnGlnGlnLysAsnGly 5944
QY 283 GTGAATGAACCCAGTCGTAATGGGTTATCTCAAGCAAGCGCAACTTTCTTAA 342
Db 5945 IleAsnAsnAlaLeuThrLysGlnLysIleGlnGlnAlaLysAla-----GlnLeuAlaGln 5963

```

Tue Mar 7 12:58:25 2006

us-09-360-685c-26.rarbm

Page 6

[illegible]

Alignment Scores:	
Pred. No.:	0 00164
Score:	13.50
Percent Similarity:	44.9%
Best Local Similarity:	20.6%
Query Match:	11.2%
DB:	5
Length:	1404
Matches:	51
Conservative:	60
Mismatches:	97
Indels:	39
Gaps:	0

US-09-360-685C-26 (1-685) X US-10-732-923-3304 (1-1404)

QY AAAATGGCAAAATATAGATTTCACAGCTAACGACAAAAGCCAGCTTGAAAT 60
Db 112 lysaenlyslgylgicigicigilnapegimlylvalginalawetgimlphelyls 133
QY 61 TCCGTTAAATGTG-----ATCATCATCTAAAGGTACCGATTAAGTTGAATCTTC 114
Db 132 SerGlnlyslValLeuGlyAspLeuileLysenSerAspAspGlyGlnmryAlaAspAsnSer 153
QY 115 AATCAACGGGTATCGAGGGCTAAACACG---GGTCAATTCTGAGGGGTAGACCAAGC 174
Db 152 AsnAlaIyGdlSerLeuGlnAsnAsnThrValaAsnAluYsSerIrrLlGdlVnIle 174

Qy 172 TTAGCCGATCTCAAAAAATTCTCAAAGCAGCAATTGCCCAACAGACTCAAAAAATGAA 23
::: :::::
Db 172 Tleghlailleserlysaltsahgnulalylsygluleundllseeglnlleamnan 19
::: :::::
Qy 232 AGTCATCATGCTGGAAAAAATCTGAATATATCATTCCTTAAGAATGTGTGAATGGA 291
::: |||
Db 192 AAlaygenhnglnuluys-----GlualPheglnluLyalygIngnlteuLeuyn 206
::: |||
Qy 292 ACCCGAGTCCGTTAAATCGGTTA-----TCTCAAGCAAGAACCAACTCTTTTAA 344
::: |||
Db 210 Leulleyserhenhnluleapnaserlyslvalapsgluthtlailtleuyn 228
::: |||
Qy 343 AACTTTTGGAGATC-----
::: |||
Db 230 Aanthrxnsvalvalalglyabserlilelysthllelguthtyrtyrlyglutle 249
::: |||
Qy 370 AATGCAAAATCGGAATTTCAATTAACAATAACATPATAGCACTCAAAAACGACCATT 429
::: |||
Db 250 Glulyalalaleulseteuthrxnsnyallleanglu--Phelyslyeglulnglu 268
::: |||
Qy 430 TATCTTAAAGTTATATAAAAAGAAACAGGCGACAGCGTAGCTTGAAAGCACCTTATAC 489
::: |||
Db 265 LyxAlaasnvalLyvalavalPheSerlyslseteryeglnteuLysapreultleap 288
::: |||
Qy 490 GCTCAAGTCTAAAAAGTTAATSCAAAAATTGACCCGCTCATCTCAATATACAAAGTGT 549
::: |||
Db 289 SerclunspgllyslrsValaspeSerfashngu----SerglnvalLeuthr- 305
::: |||
Qy 550 TTGGGTGTGTAGGGCAAGCAGCGGCTTCCCTTTGAAAAGCACTGATTAATGGATGAT 609
::: |||
Db 306 -----Lystrhytyleagpdu 311
::: |||
Qy 610 CTCAAGTAAAGTAAAGG-----CTTCAAGAGATCAAGAATGGCTCGAGAAATGACAAAT 663
::: |||
Db 312 AansSerSetleegliunspillegliuasnlythrlysaepillegluluyalalleulset 331
::: |||
Qy 664 CTCATCAAGCGGTATCAAGA 684
::: |||
Db 332 Leuthrxnslyslleamnap 338
::: |||

RESULT 8
US-10-282-122A-70177

Publication No. US20040029129A1
: GENERAL INFORMATION:

APPLICANT: Zamudio, Carlos

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari

APPLICANT:	Wall, Daniel
APPLICANT:	Trawick, John

APPLICANT: Yamamoto, Robert

TITLE OF INVENTION: Identification of E

; CURRENT APPLICATION NUMBER: US/10/282,122
 ; CURRENT FILING DATE: 2003-03-20

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/207.727

PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06

PRIOR FILING DATE: 2000-09-09

EXACT FILING DATE: 2000-10-23


```

Dd      1286  LeuluyvApyAxiIleanginIleleuInginIyVhIeAmberIleAmAmaIleu  1305
Qy      239  GTCCGTAAAGGGTTATCTCCAGAGAACCCCAACTCTTTTCTTAAAACTTTTCGACATC  357
Dd      1306  ThryIeIugInIleIugInIAlayAla---GIneulAlaGInIleuInIAspIle  1324
Qy      358  AAGAAAGAGTGAATGCAAACTTGGAATTTCAATTAACAATTAACAATTAAGACTCAAA  417
Dd      1325  Lys-----  1325
Qy      418  AAGAACCCATTTTGTGTAATTAATTAAGAAAGAGGCGACAGCTAGCTTGAA  477
Dd      1326  -AepIeValIyAlayIeIuMepIeIyVAmhAlIleIyAlIeAlaIa-----  1342
Qy      478  GAACCATTTAGCTCAAGTCTGTAATAAGATAATGCAAAAATTGACCGACTCATCA  537
Dd      1343  -----AmaIAlayvAryApGInIleAen-----  1350
Qy      538  ATAGCAAGCTTTGGGTGTGTGAGGCAACGACGGGCTTCCTTGAAAGGCATGAT  597
Dd      1351  -----SeIaAmPfoAepIeuthrProGInIyAlayAlaIeulV-----  1365
Qy      598  AAAGTGTAGTACTCAAGTAAGTAGAGGCTTTCAGAGAAATCAAGAAATTTGGCTCA  657
Dd      1366  GIuItIeApGInIleIugIyAryAlaIeU---GIneVAlaIuAmhIleGInItrIle  1384
Qy      658  GACAATTCATCAAGCGGTATCA  661
Dd      1385  AepGInIeAmhAryAlayIeulAen  1392

```

```

RESULT 10
US-10-470-048B-440
Sequence 440. Application US/10470048B
Publication No. US2005003744A1
GENERAL INFORMATION:
APPLICANT: MEINKE ET AL.
TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
FILE REFERENCE: SOUN:03505
CURRENT APPLICATION NUMBER: US/10/470,048B
CURRENT FILING DATE: 2003-07-25
NUMBER OF SEQ ID NOS: 603
SOFTWARE: PatentIn version 3.1
SEQ ID NO 440
LENGTH: 10498
TYPE: PRT
ORGANISM: Staphylococcus aureus
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (9728)..(9731)
OTHER INFORMATION: X = anything
US-10-470-048B-440

```

Alignment Scores:	
Prod. No.:	
Score:	0.00377
Score:	129.00
Percent Similarity:	39.2%
Best Local Similarity:	24.3%
Query Match:	11.0%
DB:	5
	13
Gaps:	92
Length:	10498
Matches:	65
Conservative:	40
Mismatches:	71
Indels:	92
	13

```

US-09-360-685C-26 (1-685) x US-10-470-048B-440 (1-10498)
OY      16  AAGCATTTTCAGCAGAGTAAAGCAA---GCCAAAAAGGACCTTGAATAATTCGTTAAAGAT 72
          |||      |||      |||      |||      |||      |||      |||      |||
Db      9677  LykaspIenvuIyVsalIalysgIunspIalIyGInspIalysgInuValgInaIa  9696
          |||      |||      |||      |||      |||      |||      |||      |||
OY      73  GTGATC-----ATCAATCAAAAGGTACGAGTAA-----GTT 105
          |||      |||      |||      |||      |||      |||      |||      |||
Db      9697  leuIIeaSpGIuIIeaSpGInaInpIraCnaInuItnIaSpIyGInuIyGInaIaleuIy  9716
          |||      |||      |||      |||      |||      |||      |||      |||
OY      106  GATPATCTCAATCAA-----GTT 120
          |||      |||      |||      |||      |||      |||      |||      |||

```

Db 9717 AspArgLeuGlnGlnIleuLeuGlnGlnIYhiE***AspIle***AsnIaWetThrIyS 9736

QY 121 ---CGGGTATCAAGTGGCTAAAGCAACGGGTGATTTCAAGTGGTAAAGCAACGGTATGCC 177

Db 9737 GluAlaIleuGlnAlaIyS-----GluArgLeuAlaGlnIleuGln 9751

QY 178 GATCTCAAAAATTTCTCAAG----- 198

Db 9752 AspIleIyAspIleValIySaIaIySgluAspAlaIySaEnAspIleAspIyArgVal 9771

QY 199 GAGCAATTTGGCCCAACAGCATCTCAAAAATGAAGCTCATCTGTAAGAAAAA----- 252

Db 9772 GlnAlaLeuIleAspGlnIleAspGlnIleAspGlnAsProAsnLeuThrAspIySgluYglnAla 9791

QY 253 -----TCGAATATATATCATCCGTTAAAGATGCTGTGTAATGCAACCTTA 297

Db 9792 LeuIyAspArgIleAsnGlnIleAsnGlnGlnGlnIYhiSaEnAspIleAsnAsnAlaIyS 9811

QY 298 GTCCGATATAGGTTATCTCATCAAGAAAGCCACAACTCTTTCTAAAACTTTGGACATC 357

Db 9812 ThrIySgluGlnIleGlnGlnAlaIySaIa-----GlnIyAsnIleGlnIleuGlnAspIle 9830

QY 358 AAGAAAGAGTGAATGCAAAACTTGGAATTTCAATTAACAATTAACATATGATGATCAAA 417

Db 9831 Lys----- 9831

QY 418 AACGAACCATTTATGCTAAAGTTAATTAAGAAAGACGGACAGCATGACTTGAA 477

Db 9832 ---AspLeuValIySaIaIySgluAspAlaIySaEnAlaIleIyAlaIleuAla----- 9848

QY 478 GAACCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTTGACCAACATCA 537

Db 9849 -----AsnAlaIyAspArgGlnIleAsn----- 9856

QY 538 ATAGCAAGTGTGTGGTGTGTAGGGCAACAGACGGGGCTCCCTTGAATAAGCATAT 597

Db 9857 -----SerAsnProAspLeuThrProGlnGlnIySaIaIyAlaIleuIyS----- 9871

QY 598 AAAGTGAATGATCTACAGTAAAGTATGGGCTTTCAAGATCAAGAAATGGCTCAAAAAT 657

Db 9872 GlnIleAspGlnAlaGlnIyAspArgAlaIyS-----GlnAsnValGlnAsnAlaGlnIle 9890

QY 658 GACATCTCATCAACACGGTATCA 681

Db 9891 AspGlnIleuAsnArgIlyyEnAsn 9898

RESULT 11
US-09-815-242-5893
Sequence 5893 Application US/09815242
Patent No US20020061565A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Olesen, Karl L.
APPLICANT: Zysekund, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
PILE REFERENCE: E11TRA.011A
CURRENT FILING DATE: 2001-03-21
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625

/ PRIOR FILING DATE: 2000-11-27
 / PRIOR APPLICATION NUMBER: 60/257,931
 / PRIOR FILING DATE: 2000-12-22
 / PRIOR APPLICATION NUMBER: 60/269,308
 / PRIOR FILING DATE: 2001-02-16
 / NUMBER OF SEQ ID NOS: 14110
 / SOFTWARE: FastSeq for Windows Version 4.0
 / SEQ ID NO 5883
 / LENGTH: 837
 / TYPE: PRF
 / ORGANISM: Staphylococcus aureus
 / US-09-815-242-5883

Alignment Scores:
 Pred. No.: 0.004 Length: 837
 Score: 127.00 Matches: 69
 Percent Similarity: 37.1% Conservative: 32
 Best Local Similarity: 25.4% Mismatches: 93
 Query Match: 10.8% Indels: 78
 Gaps: 12

US-09-360-685C-26 (1-685) x US-09-815-242-5883 (1-837)

QY 7 GGCAGAAATATGAGATTTCAGCAGGTACGCAAGCAAAAAGCGACCTTGAATAATTCGTT 66
 DB 470 GYAENGUGLNLySLAAsnValAspAlaMetAspLysLeuArgInSerIle 489
 QY 67 AAAGATGATCATC-----AATCAAAAGGTACGAT-----AAAGTT 105
 DB 490 AAlAspAsnAlaThrThrLysGlnAsnGlnAsnTyThrAspLaseGlnAsnLys 509
 QY 106 GATTAATCTCATCAAGCGGTATCAAGCGCT----- 135
 DB 510 AspAlaTyraAsnAsnAlaValThrThrAlaGlnGlyIleIleAspGlnThrSerPro 529
 QY 136 -----AAAGCAAGCGGTATTCAGTACGAGTACGAGCAAGCG 171
 DB 530 ThrLeuAspProThrValIleAsnGlnAlaAlaGlyGlnValSerThrThrLysAsnAla 549
 QY 172 TTAGCCGATCTCAAAAATTTTC-----TCAAGAGCAATTTGGCCCAAGCTCAAAA 225
 DB 550 LeuAsnGlyAsnGlnLysLeuGlnAlaAlaLysGlnGlnAlaSerGlnSerLeuGlySer 569
 QY 226 AATGAAGTCTCAATCTGAAAATAATCTGAATATATCATCCGTTAGATGCTG 285
 DB 570 LeuAspAsnLeuAsnAsnAlaGlnLys-----GlnThrValThrAspGlnIle 585
 QY 286 AATGGA----- 291
 DB 586 AsnGlyAlaIleThrValAspGlyAlaAsnGlnIleLysGlnAsnAlaGlnAsnLeuAsn 605
 QY 292 ACCCTAGTGGTAAAT-----GGTTATCTCAAGCAAGCAAGCAAGCTTTCTTAA 342
 DB 606 ThrAlaMetGlyAsnLeuLysGlnAlaIleAlaAspLysAspAlaThrLysAlaThrVal 625
 QY 343 AACTTTGGAGATCAAGAAAGAGTTGATGCAAAATCTGAAATTCATTAACAATAAC 402
 DB 626 AsnProThrAspAlaAspGln-----AlaLysGlnAlaIleTyraSerThrAlaVal 642
 QY 403 AATAATGAGATCAAAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 462
 DB 643 ThrAsnAla-----GluAsnIleIleSerLysAlaAsnGlyLysAlaAsnAlaThrGln 659
 QY 463 GCACTAGGCTTGAAGAACCACTTACGCTCAAGTGTCTAAAGAGTAATGCAAAAAT 522
 DB 660 AlaGlu-----ValGlnGlnAlaIleLysGlnValAsnAlaLys 673
 QY 523 GACGACTCAATCAATGACAGTGTGGTGGTGTGTAAGGCAAGCAAGCGGCTTCCCT 582
 DB 674 GlnAlaLeuAsnGlyAsnAlaAsn----- 681
 QY 583 TTGAAGAAGCATGATTAAGTTGATGATCTCAGTAAGTAGGCTTTCAAGGAATCAAGA 642
 DB 642 ----- 681

DB 682 ---ValGlnIleAlaLysAspGlnAlaThrAlaLeuIleAsnSerSerAsnAspLeuAsn 700
 QY 643 TTGGCTCAGAAAATGACAAATTCATCAATCAAGCGGTA 678
 DB 701 GlnAlaGlnLys---AspAlaLeuLysGlnGlnVal 711

RESULT 12
 US-09-815-242-13080
 / Sequence 13080, Application US/09815242
 / Patent No. US20020061569A1
 / GENERAL INFORMATION:

/ APPLICANT: Haselbeck, Robert
 / APPLICANT: Ohlsen, Kari L.
 / APPLICANT: Zyskind, Judith W.
 / APPLICANT: Wall, Daniel
 / APPLICANT: Trawick, John D.
 / APPLICANT: Carr, Grant J.
 / APPLICANT: Yamamoto, Robert T.
 / APPLICANT: Xu, H. Howard

/ TITLE OF INVENTION: Identification of Essential Genes in
 / FILE REFERENCE: ELITRA.011A
 / CURRENT APPLICATION NUMBER: US/09/815,242
 / PRIOR FILING DATE: 2001-03-21

/ PRIOR APPLICATION NUMBER: 60/191,078
 / PRIOR FILING DATE: 2000-03-21
 / PRIOR APPLICATION NUMBER: 60/206,848
 / PRIOR FILING DATE: 2000-05-23

/ PRIOR APPLICATION NUMBER: 60/207,727
 / PRIOR FILING DATE: 2000-05-26
 / PRIOR APPLICATION NUMBER: 60/242,578
 / PRIOR FILING DATE: 2000-10-23

/ PRIOR APPLICATION NUMBER: 60/253,625
 / PRIOR FILING DATE: 2000-11-27
 / PRIOR APPLICATION NUMBER: 60/257,931
 / PRIOR FILING DATE: 2000-12-22

/ PRIOR APPLICATION NUMBER: 60/269,308
 / PRIOR FILING DATE: 2001-02-16
 / NUMBER OF SEQ ID NOS: 14110
 / SOFTWARE: FastSeq for Windows Version 4.0
 / SEQ ID NO 13080

/ LENGTH: 875
 / TYPE: PRF
 / ORGANISM: Staphylococcus aureus
 / US-09-815-242-13080

Alignment Scores:
 Pred. No.: 0.00403 Length: 875
 Score: 127.00 Matches: 69
 Percent Similarity: 37.1% Conservative: 32
 Best Local Similarity: 25.4% Mismatches: 93
 Query Match: 10.8% Indels: 78
 Gaps: 12

US-09-360-685C-26 (1-685) x US-09-815-242-13080 (1-875)
 QY 7 GGCAGAAATATGAGATTTCAGCAGGTACGCAAGCAAAAAGCGACCTTGAATAATTCGTT 66
 DB 507 GYAENGUGLNLySLAAsnValAspAlaMetAspLysLeuArgInSerIle 526
 QY 67 AAAGATGATCATC-----AATCAAAAGGTACGAT-----AAAGTT 105
 DB 527 AAlAspAsnAlaThrThrLysGlnAsnGlnAsnTyThrAspLaseGlnAsnLys 546
 QY 106 GATTAATCTCATCAAGCGGTATCAAGCGCT----- 135
 DB 547 AspAlaTyraAsnAsnAlaValThrThrAlaGlnGlyIleIleAspGlnThrSerPro 566
 QY 136 -----AAAGCAAGCGGTATTCAGTACGAGTACGAGCAAGCG 171
 DB 567 ThrLeuAspProThrValIleAsnGlnAlaAlaGlyGlnValSerThrThrLysAsnAla 586
 QY 172 TTAGCCGATCTCAAAAATTTTC-----TCAAGAGCAATTTGGCCCAAGCTCAAAA 225

Tue Mar 7 12:58:25 2006

us-09-360-685c-26.rapbm

Page 12

[illegible]

Search completed: March 6, 2006, 20:22:24
Job time : 176.5 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 6, 2006, 20:00:26 ; Search time 7.1 Seconds
(without alignments)
1595.291 Million cell updates/sec

Title: US-09-360-685C-26
Perfect score: 1171
Sequence: 1 aaaaatgcaataaaga.....caatcaagcgatcagaag 685

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Lasting first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/abs/ABSWEB.spool/US0936085/runat.06032006.121320.15444/app.query.fasta_1
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-HOST=abgs02p -USER=US0936085@CGN_1_71@runat.06032006.121320.15444
-NCPU=6 -ICPU=3 -NO_MMAP -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*

- 1: /cgn2_6/prodata/1/1aa/5-COMB.pep:*
- 2: /cgn2_6/prodata/1/1aa/6-COMB.pep:*
- 3: /cgn2_6/prodata/1/1aa/H-COMB.pep:*
- 4: /cgn2_6/prodata/1/1aa/CTUS-COMB.pep:*
- 5: /cgn2_6/prodata/1/1aa/RE-COMB.pep:*
- 6: /cgn2_6/prodata/1/1aa/backfill1est.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1127	96.2	1147	2	US-08-470-260-5
2	1127	96.2	1147	2	US-08-471-491-5
3	1127	96.2	1147	2	US-08-466-662-5
4	1127	96.2	1147	2	US-08-477-451-2
5	1081	92.3	3200	1	US-08-477-451-8
6	1030.5	88.0	268	2	US-08-993-010-4
7	949	81.0	3287	1	US-08-477-451-7
8	930	79.4	3169	1	US-08-477-451-6
9	927.5	79.2	3174	1	US-08-477-451-3
10	902.5	77.1	1181	1	US-08-053-614-4
11	902.5	77.1	1181	1	US-08-316-397B-4
12	902.5	77.1	1181	1	US-08-034-306-4

13	902.5	77.1	1181	2	US-09-259-437-4	Sequence 4, Appli
14	902.5	77.1	1181	4	PCT-US93-09782-4	Sequence 4, Appli
15	798.5	68.2	3177	1	US-08-477-451-4	Sequence 4, Appli
16	454.5	38.8	859	1	US-08-053-614-2	Sequence 2, Appli
17	454.5	38.8	859	1	US-08-316-397B-2	Sequence 2, Appli
18	454.5	38.8	859	1	US-09-034-306-2	Sequence 2, Appli
19	454.5	38.8	859	2	US-09-259-437-2	Sequence 2, Appli
20	454.5	38.8	859	4	PCT-US93-09782-2	Sequence 2, Appli
21	123	10.5	10182	2	US-09-134-001C-3159	Sequence 3159, Ap
22	120.5	10.3	5024	2	US-09-710-279-2564	Sequence 2964, Ap
23	118	10.1	1693	2	US-09-560-385A-4	Sequence 4, Appli
24	118	10.1	1693	2	US-09-560-385A-8	Sequence 8, Appli
25	118	10.1	1713	2	US-08-600-982-24	Sequence 24, Appli
26	118	10.1	1713	2	US-09-560-385A-5	Sequence 6, Appli
27	118	10.1	1713	2	US-09-538-092-1359	Sequence 1359, Ap
28	118	10.1	1713	4	PCT-US94-10261A-24	Sequence 24, Appli
29	118	10.1	1724	2	US-09-560-385A-2	Sequence 2, Appli
30	117	10.0	693	2	US-08-235-836C-68	Sequence 68, Appli
31	115.5	9.9	1095	2	US-09-710-279-3154	Sequence 3154, Ap
32	115	9.8	873	2	US-09-336-447A-13	Sequence 13, Appli
33	115	9.8	873	2	US-09-952-267B-13	Sequence 13, Appli
34	114	9.7	3712	2	US-10-037-417-48	Sequence 48, Appli
35	113.5	9.7	1401	2	US-09-750-590A-2	Sequence 2, Appli
36	113.5	9.7	1939	2	US-09-310-187A-1	Sequence 1, Appli
37	113.5	9.7	1939	2	US-09-538-092-917	Sequence 917, Appli
38	113	9.6	878	2	US-09-540-236-3401	Sequence 3401, Ap
39	113	9.6	3696	2	US-09-134-001C-5080	Sequence 5080, Ap
40	112.5	9.6	611	2	US-09-710-279-2586	Sequence 2586, Ap
41	112.5	9.6	718	2	US-09-540-236-2753	Sequence 2753, Ap
42	112	9.6	1935	2	US-09-538-092-916	Sequence 916, Appli
43	112	9.6	3712	2	US-10-037-417-51	Sequence 51, Appli
44	111.5	9.5	700	1	US-07-720-588-2	Sequence 2, Appli
45	111.5	9.5	700	1	US-08-785-190-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-470-260-5
; Sequence 5, Application US/08470260
; Patent No. 6077706
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,260
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/256,848
; FILING DATE: 21-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0316.001

Tue Mar 7 12:58:24 2006

US-09-360-685C-26.ra1

Page 2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1147 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-260-5

Alignment Scores:
Pred. No.: 5,536-108 Length: 1147
Score: 1127.00 Matches: 228
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.2% Indels: 0
DB: 2 Gaps: 0

US-09-360-685C-26 (1-685) x US-08-470-260-5 (1-1147)

QY 1 AAAAATGCAAAAATTAAGATTTCAGCAAGTAACGCAAAAAGCACTTGAAT 60
Db 750 LysaennglylysaenlysaepPheSerlyValThrGlnAlaLysSerAspLeuGln 769
QY 61 TCCGTTAAAGATGATCATCATCAAAAGTTACGGATTAAGTTGATTCATCA 120
Db 770 SerVallysaepValIleIleAenGlnlyValThrAspLysValAspAsnLeu 789
QY 121 GCGGATCAGTGGCTAAAGCAAGCGGTGATTTCAGTAAGGTAGCAAGCTTAC 180
Db 790 AlaValSerValAlaLysAlaThrGlnLysPheSerAspValGlnGlnAlaLeuAla 809
QY 181 CTCAAAAATTTCTCAAGAGCAAGATTGGCCCAACAGCTCAAAAAATGAAAGTCT 240
Db 810 LeuLysaenPheSerlyGlnGlnLeuAlaGlnGlnAlaGlnLysaenGlnSerLeu 829
QY 241 GGTAAATGATGATCAATCAATCCGTTAAGATGATGATGATGATGATGATGAT 300
Db 830 AlaValSerValAlaLysAlaThrGlnLysPheSerAspValGlnGlnAlaLeuAla 809
QY 301 GGTAAATGATGATCAATCAATCCGTTAAGATGATGATGATGATGATGATGAT 360
Db 850 GlnAenGlnLysSerGlnAlaGlnAlaThrLysSerLysaenPheSerAspIleLys 869
QY 361 AAAGAGTTGAATGCAAACTTGAAATTTCAATTAACAATTAACAATTAACAAT 420
Db 870 LysGlnLysaenAlaLysLeuGlnLysaenPheAsnAsnAsnAsnAsnGlnLysaen 889
QY 421 GAACCAATTTATGCTTAAGTTAATTAAGAAAGCAAGCGCAAGCTGACCTGAAGA 480
Db 890 GlnProIleTyraLysValAlaLysValLysValAlaGlnAlaLysSerLeuGln 909
QY 481 CCCATTACGCTCAAGTGTCTTAAGTTAAGTTAAGTTAAGTTAAGTTAAGTTAAG 540
Db 910 ProIleTyraLysValAlaLysValLysValAlaLysValLysValLysValLys 929
QY 541 GCAAGTGTGGGTGTGTAGGCAAGCGGCTTCCCTTGAAGAAAGCAATGAATAA 600
Db 930 AlaSerGlnLeuGlnValAlaGlnAlaLysValLysValLysValLysValLysVal 949
QY 601 GTTGATGATTCAGTAAGTGAAGGCTTTCAAGAAATCAAAATGGCTCAAAATTTGAC 660
Db 950 ValAspAspLeuSerlyValGlnLysSerLysaenGlnLysaenGlnLysaen 969
QY 661 AATCTCAATCAAGCGGTATGAGA 684
Db 970 AsnLeuAenGlnAlaValSerGln 977

RESULT 2
US-08-471-491-5
Sequence 5, Application US/08471491B

Patent No. 6090611
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
APPLICANT: Burgoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappoli, Rino
TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
FILE REFERENCE: CHIR0044
CURRENT APPLICATION NUMBER: US/08/471.491B
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 1147
TYPE: PRT
ORGANISM: Helicobacter pylori
US-08-471-491-5

Alignment Scores:
Pred. No.: 5,536-108 Length: 1147
Score: 1127.00 Matches: 228
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.2% Indels: 0
DB: 2 Gaps: 0

US-09-360-685C-26 (1-685) x US-08-471-491-5 (1-1147)

QY 1 AAAAATGCAAAAATTAAGATTTCAGCAAGTAACGCAAAAAGCACTTGAAT 60
Db 750 LysaennglylysaenlysaepPheSerlyValThrGlnAlaLysSerAspLeuGln 769
QY 61 TCCGTTAAAGATGATCATCATCAAAAGTTACGGATTAAGTTGATTCATCA 120
Db 770 SerVallysaepValIleIleAenGlnlyValThrAspLysValAspAsnLeu 789
QY 121 GCGGATCAGTGGCTAAAGCAAGCGGTGATTTCAGTAAGGTAGCAAGCTTAC 180
Db 790 AlaValSerValAlaLysAlaThrGlnLysPheSerAspValGlnGlnAlaLeuAla 809
QY 181 CTCAAAAATTTCTCAAGAGCAAGATTGGCCCAACAGCTCAAAAAATGAAAGTCT 240
Db 810 LeuLysaenPheSerlyGlnGlnLeuAlaGlnGlnAlaGlnLysaenGlnSerLeu 829
QY 241 GGTAAATGATGATCAATCAATCCGTTAAGATGATGATGATGATGATGATGAT 300
Db 830 AlaValSerValAlaLysAlaThrGlnLysPheSerAspValGlnGlnAlaLeuAla 809
QY 301 GGTAAATGATGATCAATCAATCCGTTAAGATGATGATGATGATGATGATGAT 360
Db 850 GlnAenGlnLysSerGlnAlaGlnAlaThrLysSerLysaenPheSerAspIleLys 869
QY 361 AAAGAGTTGAATGCAAACTTGAAATTTCAATTAACAATTAACAATTAACAAT 420
Db 870 LysGlnLysaenAlaLysLeuGlnLysaenPheAsnAsnAsnAsnAsnGlnLysaen 889
QY 421 GAACCAATTTATGCTTAAGTTAATTAAGAAAGCAAGCGCAAGCTGACCTGAAGA 480
Db 890 GlnProIleTyraLysValAlaLysValLysValAlaGlnAlaLysSerLeuGln 909
QY 481 CCCATTACGCTCAAGTGTCTTAAGTTAAGTTAAGTTAAGTTAAGTTAAGTTAAG 540
Db 910 ProIleTyraLysValAlaLysValLysValAlaLysValLysValLysValLys 929
QY 541 GCAAGTGTGGGTGTGTAGGCAAGCGGCTTCCCTTGAAGAAAGCAATGAATAA 600
Db 930 AlaSerGlnLeuGlnValAlaGlnAlaLysValLysValLysValLysValLysVal 949
QY 601 GTTGATGATTCAGTAAGTGAAGGCTTTCAAGAAATCAAAATGGCTCAAAATTTGAC 660
Db 950 ValAspAspLeuSerlyValGlnLysSerLysaenGlnLysaenGlnLysaen 969

```

OY      661  AATCTCAATCAAGGGATACAGA 684
      |||
DB      970  AsnLeuasnGlnAlaValIserGlu 977

RESULT 3
US-08-466-662-5
; Sequence 5, Application US/08466662B
; Patent No. 6130059
; GENERAL INFORMATION
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappunoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; FILE REFERENCE: CHIR0057
; CURRENT APPLICATION NUMBER: US/08/466,662B
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1147
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; US-08-466-662-5

```

```

Alignment Scores:
Pred. No.:      5,53e-108      length:      1147
Score:          1137.00      Matches:      228
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches:  0
Query Match:    96.2%      Indels:      0
DB:             2          Gaps:             0

US-09-360-685C-26 (1-685) x US-08-466-662-5 (1-1147)

```

QY	AAAAATGGCAAAAATAAGGATTTGACGACAGTAAACGCAAGCAAAAAGCACTTGAAAT	60
Db	750 LysAsnGlyLysAsnLysAspPheSerLysValThrGlnAlaLysSerAspLeuLysn	765
QY	61 TCCGTTAAAGATGTGATCAATCAATCAAAAGGTAAACGATAAAGTTGATATCATCA	120
Db	770 SerValLysAspValIleIleLeuGlnLysValThrAspLysValAspAsnLeuAsnGln	785
QY	121 GCGGTATCAATGGCTAAAGCAACGGGTGATTTCAAGTAGAGACCAAGCGTTAGCGAT	180
Db	790 AlaValSerValAlaLysAlaThrGlyAspPheSerArgValGluGlnAlaLeuAlaAsp	805
QY	181 CTCAAAAATTTCTCAAGAGCAAAATTGGCCCAACAGCTCAAAAAATGAAAGTCTCAAT	240
Db	810 LeuLysAsnPheserLysGluGlnLeuAlaGlnGlnAlaGlnLysAsnGlyLysLeuAsn	825
QY	241 GCTAGAAAAAAATCTGAATATATCAATCCGTTAAGAAATGGTGTAATGAAACCTTAGTC	300
Db	830 AlaArgLysLysSerGluIleTyrGlnSerValLysAsnGlyLysAsnGlyThrLeuVal	845
QY	301 GGTAAATGGGTTATCTCAACAGCAAGCAACATCTTTCTAAAAAATTTTCGACATCAAG	360
Db	850 GlyAsnGlyLysSerGlnAlaGluAlaThrThrLeuSerLysAsnPheserAspIleLys	865
QY	361 AAAGAGTTGAATGCAAAACTTGGAATTTCAATTAACAATTAACAATATGACTCAAAAAC	420
Db	870 LysGluLeuAsnAlaLysLeuGlyAsnPheAsnAsnAsnAsnAsnGlyLeuLysAsn	885
QY	421 GAACCAATTATGCTAAAGTTAATATAAAGAAAGAGGGCAAGCGTACGCTTGAAAGAA	480
Db	890 GluProIleTyrAlaLysValAsnLysLysLysLysAlaGlyGlnAlaAlaSerLeuGluGlu	905
QY	481 CCCATTACAGCTCAACTTGCTAAAAAAGGTAAATGCAAAAATTTGACCGCATCAATCAATA	540
Db	910 ProIleTyrAlaGlnValAlaLysLysValAsnAlaLysIleAspArgLeuAsnGlnIle	925

QY 541 GCAAGTGGTTTGGGTGTGTGTAGCGCAGACGCGGCTTCCCTTTGAAGAAGCATGATMAA 600
Db 930 AlaseRGlyleuGlyValValAlGlylnAlaIaGlyPheProleuLysValGHIaBpRlye 949
QY 601 GTTGATGATCTCAATNAGTAGAGGCTTTCAAGAAATCAAGAAATGGCTCAAGAAATTTGAC 660
Db 950 ValAspAspLeuSerLysValAlGlyLeuSerArgAsnGlnGluLeuAlaGlnLysIleAsp 969
QY 661 AATCTCAATCAAGCGGTATCAGAA 684
Db 970 AsnLeuAsnGlnAlaValAlaSerGln 977

```

1 RESULT 4
2 US-08-477-451-2
3 Sequence 2, Application US/08477451
4 Patent No. 592865
5
6 GENERAL INFORMATION:
7 APPLICANT: Covacci, Antonello
8 TITLE OF INVENTION: Helicobacter Pylori CagI Region
9 NUMBER OF SEQUENCES: 46
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: Chiron Corporation
12 STREET: 4560 Horton Street
13 CITY: Emeryville
14 STATE: CA
15 COUNTRY: USA
16 ZIP: 94608-2916
17
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC-DOS/MS-DOS
22 SOFTWARE: Patentin Release #1.0, Version #1.30
23
24 CURRENT APPLICATION DATA:
25 APPLICATION NUMBER: US/08/477,451
26 FILING DATE: 07-JUN-1995
27 CLASSIFICATION: 435
28
29 ATTORNEY/AGENT INFORMATION:
30 NAME: McClung, Barbara G.
31 REGISTRATION NUMBER: 33,113
32 REFERENCE/DOCKET NUMBER: 035.002
33
34 TELECOMMUNICATION INFORMATION:
35 TELEPHONE: 510-601-2708
36 TELEFAX: 510-655-3542
37
38 INFORMATION FOR SEQ ID NO: 2:
39 SEQUENCE CHARACTERISTICS:
40 LENGTH: 3289 amino acids
41 TYPE: amino acid
42 STRANDEDNESS: single
43 TOPOLOGY: linear
44
45 MOLECULE TYPE: protein
46
47 US-08-477-451-2

```

Alignment Scores:	
Pred. No.:	7,476-108
Score:	1127.00
Percent Similarity:	100.0%
Best Local Similarity:	100.0%
Query Match:	96.2%
DB:	1
Gaps:	0
Length:	3289
Matches:	228
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-360-685C-26 (1-685) x US-08-477-451-2 (1-3289)

QY 1 AAAATGGCAAAATAATAGATTTTCAGACAGGTACCGACGAAAAAGCAGCTTGAAAT 60

Db 2281 LysaasnGlyLysaasnLysaapPheSerLysValThhGlnAlaLysSerAspLeuGlnuSn 2300

QY 61 TCCGTTAAAGATGTGATCATCAATCAAAGGTACGATTAAGTTGATATTCATCAATA 120

Db 2301 SerValLysaapValIleIleasnGlnLysValThhAspLysValAspasnLeuasnGln 2320

QY 121 GCGGTATCATGTGGCTAAAGCAACGGGTGATTTTCAGTGGGTAGACCAAGCTTGACCGAT 180

Tue Mar 7 12:58:24 2006

US-09-360-685c-26.ra1

Page 4

Db 2321 AlaValSerValAlaValSerValThrGlyAspPheSerArgValGluGlnAlaLeuAlaAsp 2340
QY 181 CTCAAAATTTCTCAAAAGGACGATTTGGCCCAACGCTCAAAATTAATTAATGCTCAAT 240
Db 2341 LeuValAsnPheSerArgGluGlnLeuAlaGlnGlnAlaGlnValAsnGlnSerLeuVal 2360
QY 241 GCTAGAAAAAATCTGAATATATCAATCCCTTAGAAATGCTGAATGAAACCTTAGTC 300
Db 2361 AlaArgLysLysSerGluLeuValGlnSerValLysAsnGlyValAsnGlyThrLeuVal 2380
QY 301 GGTAAATGGTTATCTCAAGCAGAACCCACACTCTTTCTTAAATCTTTCCAGACTCAAG 360
Db 2381 GlyAsnGlyLeuSerGlnAlaGlnAlaThrThrLeuSerLysAsnPheSerArgPheLys 2400
QY 361 AAAGAGTTGAATGCAAACTTGAAATTTCAATTAACATTAATGAATGAATCAAAAC 420
Db 2401 LysGluLeuAsnAlaLysLeuGlyAsnPheAsnAsnAsnAlaGlnLysValAsn 2420
QY 421 GAACCAATTTATGCTTAAGTTAATTAATAAGAAAGAGGAGCAAGCTTAGCAAGAA 480
Db 2421 GluProLeuThrAlaLysValAsnLysLysValAlaGlyGlnAlaLysSerLeuGln 2440
QY 481 CCATTTAGCTCAAGCTGCTTAAATAAGTAAATGCAAAATTTGACCACTCAATCAATA 540
Db 2441 ProLeuThrAlaGlnAlaLysValAsnLysValAsnAlaLysLysLeuAsnGln 2460
QY 541 GCAAGCTGTTGGCTTTGTAAGGAGCAAGAGCGGCTTCCCTTGAAGAGCAATGATAA 600
Db 2461 AlaSerGlyLeuGlyValAlaGlyGlnAlaAlaGlyPheProLeuLysArgHisAspLys 2480
QY 601 GTTATGATCTCAAGTAAGGAGCTTTCAAGCAATCAAGATTTGGCTCAAGAAATTCAG 660
Db 2481 ValAspSerLeuSerLysValGlyLeuSerArgAsnGlnLeuAlaGlnLysLeuAsp 2500
QY 661 AATCTCAATCAAGCGGTATCAGAA 684
Db 2501 AsnLeuAsnGlnAlaValSerGlu 2508

RESULT 5
US-08-477-451-8
Sequence 8, Application US/08477451
Patent No. 5928865
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter Pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-655-3542
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 3200 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-451-8

Alignment Scores:
Pred. No.: 4,22e-103
Score: 1081.00
Percent Similarity: 97.3%
Best Local Similarity: 97.3%
Query Match: 92.3%
DB: 1
Gaps: 6
Length: 3200
Matches: 220
Conservative: 0
Mismatch: 0
Indels: 6

US-09-360-685c-26 (1-685) x US-08-477-451-8 (1-3200)

QY 678 TACCGCTGATTGAGATGTCATATTTCTGAGCAATTTCTGATTCCTGAAGCCTTAC 619
Db 768 TyrArgLeuIleGluIleValAlaAsnPheLeuSerGlnPheLeuIlePro---LysProTyr 786
QY 618 CTTACGATGATCAATCAATTTATCATCCTTTTCAAGGAAAGCCGCTGCTCCCTTAC 559
Db 787 LeuThrGluIleIleAsnPheIleMetProPheGlnArgGlnAlaArgCysLeuProTyr 806
QY 558 AACACCAACCACTGCTATTTGATTTGATGCTGCTCAATTTTTCATTTACCTTTTAAAC 499
Db 807 AsnThrGlnThrThrCysArgLeuIleGluSerValAsnPheCysIleTyrLeuPheSer 826
QY 498 AACTGAGCGTAATGGGTTCTTCAAGGCTACTGCTTCCCTGCTTCTTTTATTAAAC 439
Db 827 AsnLeuSerValAsnGlyPhePheLysAlaSerCysLeuProCysPheLeuPheIleAsn 846
QY 438 TTGAGCAATTAATGGGTTCTTGAAGTCCATTTATTTGTTATTTGAATTTTCAAG 379
Db 847 PheSerIleAsnGlyPheValPheGluSerIleIleValIleGluIleSerLys 866
QY 378 TTGACATTCATCTTTCTTGAATGTCGAAAGTTTGAAGAGTTGGCTTCTGAC 319
Db 867 PheCysIleGlnLeuPheLeuAspValAlaArgValAlaPheArgLysSerCysGlyPheCys 886
QY 318 TTGAGTAACCAATTTACCGACTAGGCTTCAATTCACCAATTTCAAGATTTATAT 259
Db 887 LeuArg---ProIleHisAsp---GlySerIleIleThrIleLeuAsnGlyLeuIleTyr 904
QY 258 TTGAGATTTTCTTACCATTTGAGACTTTCAATTTTGTAGCTTTGGCCCAATTTGCTC 139
Db 905 PheArgPhePheSerSerIleGlnThrPheIlePheLeuSerLeuLeuGlnLeuLeu 924
QY 138 CTTTGAAGAAATTTTGAATCGCTTACGCTTCAAGCTTCTCAACCCCTGAATAATCCGTTGC 139
Db 925 Leu---GluIlePheGlnIleGly---ArgLeuLeuTyrProThrGluIleThrArgCys 942
QY 138 TTGACCACTGATACCGCTTGAATGATATATCAATTTATCCGTTATCCCTTGAATGAT 79
Db 943 PheSerHis---TyrArgLeuIleGluIleIleAsnPheIleArgLysLeuLeuLeuAsp 961
QY 78 GATCACTCTTTAAGCAATTTTCAAGTGGCTTTTCTGCTTCTGCTTCACTTCTGGAATC 19
Db 962 AspHisIlePheAsnGlyIlePheLysValAlaPheCysLeuArgTyrLeuAlaGluIle 981
QY 18 CTTATTTTGGCAATTTT 1
Db 982 LeuIlePheAlaIlePhe 987

RESULT 6
US-08-993-010-4
Sequence 4, Application US/08993010
Patent No. 6902903
GENERAL INFORMATION:
APPLICANT: Quan, Stella
APPLICANT: Valenzuela, Pablo
TITLE OF INVENTION: HELICOBACTER PYLORI DIAGNOSTICS
FILE REFERENCE: 1360.002

CURRENT APPLICATION NUMBER: US/08/993, 010
CURRENT FILING DATE: 1997-12-18
EARLIER APPLICATION NUMBER: 60/033, 707
EARLIER FILING DATE: 1996-12-19
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 268
TYPE: PRF
ORGANISM: Helicobacter pylori
US-08-993-010-4

Alignment Scores:
Pred. No.: 3,45e-98 Length: 268
Score: 1030.50 Matches: 217
Percent Similarity: 83.4% Conservative: 4
Best Local Similarity: 81.9% Mismatches: 7
Query Match: 88.0% Indels: 37
Gaps: 3
DB:

US-09-360-685C-26 (1-685) x US-08-993-010-4 (1-268)

QY 1 AAAAATGCGAAAAATAGAGATTTCAGCAAGTAAAGCAAGCAAAAAGCACTTGAATAAT 60
DB 4 LysAsnGlyLysAsnLysAspPheSerLysValThrGlnAlaLysSerAspLeuGln 23
QY 61 TCCGTTAAGATGTCATCATCAAAAAGTAAAGCAATTAAGTTGATTAATCTCAATCAA 120
DB 24 SerLysAspValIlePheAsnGlnLysIleThrAspLysValAspAspLeuAsnGln 43
QY 121 GCGGTATCATGCGCTAAAGCAAGCGGTGATTTCAATAGGATAGCAAGCGGTAGCGCAT 180
DB 44 AlaValSerValAlaLysAlaThrGlyAspPheSerLysValGlnGlnAlaLeuAlaAsp 63
QY 181 CTCAAAAATTTCTCAAGAGCAATGCGCCCAAGCAAGCTCAAAAATGAAAGTCTCAAT 240
DB 64 LeuLysAsnPheSerLysGlnGlnLeuAlaGlnGlnAlaGlnLysAsnGlnSerLeuAsn 83
QY 241 GGTAGAAAAAATCTGAATATATTCATCCGTTAAGATGGTGTGATGAACCTTAGTC 300
DB 84 AlaGlyLysLysSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 103
QY 301 GGTAAATGGGTATCTCAAGAGCAAGCAAGCTCTTCAAAAAGCTTTGCAATCAAG 360
DB 104 GlyAsnGlyLysSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 123
QY 361 AAAAGATTGATGCAAAAAGCTT---GGAATTTCAATTAACAATTAACAATTAAGTCAAA 417
DB 124 LysGlnLeuAsnAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 143
QY 418 AAC-----GAACCATTTATGCTAAAGTTAATAAAGAAAGCAAGCGGCAAGCTAGC 471
DB 144 AsnSerThrGlnProLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 163
QY 472 CTGTAAGAACCACTTAAAGCTCAAGTGTCTAAAGAAAGTAAAGCAAAAATTAAGCAAGC 531
DB 164 ProGlnGlnProLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 183
QY 532 AATCAAAATGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 576
DB 184 AsnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 203
QY 576 ----- 576
DB 204 HisAspLysValAspAspLysSerLysValGlyArgSerValSerProGlnProLysLys 223
QY 577 -----TTCCTTTGAAAGCAAGCAATTAAGTGAATGAT 609
DB 224 AlaThrLysAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 243
QY 610 CTCAGTAAGTGAAGCTTCAAGCAATCAAGATGCTCAAGAAATTAATCAATCTCAAT 669
DB 244 LeuSerLysValGlyLysSerLysArgAsnGlnGlnLysLysLysLysLysLysLysLysLys 263

QY 670 CAAGCGTATCAAGAA 684
DB 264 GlnAlaValSerGln 268

RESULT 7
US-08-477-451-7
Sequence 7, Application US/08477451
Patent No. 5928865

GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA

ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335, 002

TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3287 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-451-7

Alignment Scores:
Pred. No.: 1,88e-89 Length: 3287
Score: 949.00 Matches: 214
Percent Similarity: 94.3% Conservative: 0
Best Local Similarity: 94.3% Mismatches: 1
Query Match: 81.0% Indels: 12
Gaps: 11
DB:

US-09-360-685C-26 (1-685) x US-08-477-451-7 (1-3287)

QY 663 TCTGATACCGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 624
DB 787 SerAspThrAla---LeuArgLysSerLysPhe---AlaAsnSer---PheLeuGlnSer 803
QY 623 CCTACCTTACGATGATCATCACTTATATCATGCTTTTCAAGAGAACCGGCTGCTTGC 564
DB 804 ProThrLeuLeuAspSerThrLysSerCysLeuPheLysGlyLysProAlaAlaCys 823
QY 563 CTTACACACCAACCACTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 504
DB 824 ProThrThrProLysProLeuAlaIle---LeuSerArgSerLysPheAlaPheThrPhe 842
QY 503 TTAGCACTTGAAGCGTAAATGAGTTCTTCAAGGCTAGCTGCTTGCCTTCTTTTAA 444
DB 843 LeuAlaThr-----AlaMetGlySerSerArgLeuAlaAlaCysProAlaPhePheLeu 860
QY 443 TTAACCTTACGATTAATGAGTTCGTTTGTGATGATGATGATGATGATGATGATGATGAT 384

Db	861	LeuThrIeuAla---MeCylSerPheIeuSerProIeuLeuLeuLeuLeuIuysPhe	879
Qy	383	CCAAgTTTTGCATCAACTCTTTCTTGAGATGCCAAAAGTTTTTGAAAGATGGTGCT	324
Db	880	ProSerPhehIAPhAsnSerPheIeuMetSerGluIuysPheLeuGluAlaValAla	889
Qy	323	TCGTCTGAGAAATACCCATTACCGAATAGGGTTGCATTCAACCACTTTAAACGATTGA	264
Db	900	SerAla---AspAsnProIeuProIuysAlaProIeuIuysProPheIeuIuysPhe---	917
Qy	263	TATATTCAGATTTTTTTTCATGATGATGAGACTTCATTTTTTTAGCTGTGGCCCAAT	204
Db	918	TyrIleSerPhePhePheIeuAlaIeuAluIuysSerPhePhe--AlaCysTrpAlaAsn	936
Qy	203	TGTCCTTTGAGAAATTTTTGAGATCGCTTAAGCTTGCTCACTCACTGAATCAACC	144
Db	937	CysSerPheGluIuysPheIeuAluIuysSerAlaAsnAlaCysSerTruIeuIuysSerPro	956
Qy	143	GTGGCTTTTGCCACATATACCGCTTGATTGAGATTAACAATTTATCCGTTACTTTTGA	84
Db	957	ValAlaIeuAlaIuysIuysPheTrpAla---LeuAluIuysSerTruIeuSerValIuysPhe---	974
Qy	83	TTGATATACATCTTTTAACGAAATTTCAAGGCGCTTTTGGTTGCTGATCTGCTG	24
Db	975	LeuMetIleIuysIuysSerLeuIuysIuysPheIuysSerIuysPheAlaCysValTruIeu	994
Qy	23	AAATCGTATATTTTGGCATTT 3	
Db	995	IuysSerLeuIuysPheIuysPhe 1001	

```

      RESULT 8
US-08-477-451-6
; Sequence 6, Application US/08477451
; Patent No. 5928865
      GENERAL INFORMATION:
      APPLICANT: Covacci, Antonello
      TITLE OF INVENTION: Helicobacter Pylori CagI Region
      NUMBER OF SEQUENCES: 46
      CORRESPONDENCE ADDRESS:
      ADDRESSEE: Chiron Corporation
      STREET: 4650 Horton Street
      CITY: Emeryville
      STATE: CA
      COUNTRY: USA
      ZIP: 94608-2916
      COMPUTER READABLE FORM:
      MEDIUM TYPE: floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.30
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/477,451
      FILING DATE: 07-JUN-1995
      CLASSIFICATION: 435
      ATTORNEY/AGENT INFORMATION:
      NAME: McClung, Barbara G.
      REGISTRATION NUMBER: 33,113
      REFERENCE/DOCKET NUMBER: 0335.002
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 510-601-2708
      TELEFAX: 510-655-3542
      INFORMATION FOR SEQ ID NO: 6:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 3169 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: protein
      ; US-08-477-451-6

```

Alignment Scores:	
Pred. No.:	1.71e-87
Score:	930.00
Length:	3165
Matches:	205

Percent Similarity:	89.9%	Conservative:	0
Best Local Similarity:	89.9%	Mismatches:	1
Query Match:	79.4%	Indels:	22
DB:	1	Gaps:	20

QY	685	CTTTCGATACCGGCTGATTGAGTAAATGTCATTTTTTCGAGCCAAATTCCTGATTCCTGAAA	62
Db	754	LeuLeuLeuProLeuAsp---AapCyseGlnPheSerGluProIleuLeuAspSerLeuLys	77
QY	625	GCCGACCTTACGATGAGATCATCACTTTATACAGCCCTTTTCAAGGGAGCCGGCTGCTT	56
Db	773	AlaLeuProIuYr---AapHisGlnLeuIuYrHisAlaPheSerLysGluSerProLeuLeu	79
QY	565	GCCGACACACCCAAACAACCTTGCAATTTGATGAGCGGGTCAATTTTTCATTTACCT	50
Db	792	AlaLeuGlnHisProLeuHisLeuLeuPheAsp---ValGluPheLeuHisLeuPro	81
QY	505	TTTTCAGCACTTGACCGTAAATGGGCTTTTCAGAGCTGACGCTGCTGCTGCTGCTTTT	44
Db	811	Phe---GlnLeuGlnArgLysTrpValLeuGlnGly---LeuLeuAlaLeuLeuSerPhe	82
QY	445	TATTACTTTACGATTAATGGGCTGCTTTTTCAGTTCATTTGTTATTTGTTATTAAT	38
Db	829	IuYr-----LeuHisLysTrpValArgPhe---ValHisLysTrpCyArgLysIuYr---Aan	84
QY	385	TTTCCAGTTTTCGATGATCACTGCTTTCTTATGTCGCGAAAGTTTTCAGAAAGATGTGG	32
Db	845	PheGlnValLeuHisSerTrpLeuSer---CyAspLysSerPhe---LysGluLeuTrp	86
QY	325	CTTTCGCTGGAATATACCCCTTATCCGACTGAGGGTTCATTCACACCATTTCTTAACGATT	26
Db	863	LeuLeuLeuGlnLysLeuHisLysTrpArgLeuGlyPheHisSerHisHisSer---ArgIle	88
QY	265	GATTATATTCAGATTTTTTTCTAGCATTGAGACTTCATATTTTTTTCAGCTTGTTGGCCA	20
Db	882	AapLysLeuGlnLysLeuPhe---His---AapPheHisPhePheGlnLeuValGlyPro	89
QY	205	ATTTCGCTTTGAGAAAATTTTTGAGATCGGCTTACCGCTTGCCTTACCTTACGAAATCAC	14
Db	900	IleLeuProLeuHisArgAsnPhe---AAspArgLeuTrpLeuAlaLeuProIuYr---AanHis	91
QY	145	CGCGTCCTTTAGCACTGATACCGCTTGATTGAGATTTACAACTTTATCCGTTACCTTTT	8
Db	918	ProLeuLeu---ProLeuLysProLeuAsp---AapLysTrpLeuLysIuYrProLeuProPhe	93
QY	85	GATTGATATCAATCTTTAACGGAATTTTCAAGGTGGCTTTTTCCTGCGTACCTTGC	26
Db	936	Asp-----SerHisLeu---ArgAsnPheGlnGlnLysArgPheLeuLeuAlaLeuProCys	95
QY	25	TGAATTCCTTATTTTTCGCAATTTT	2
Db	953	---AanProIuYrPheCyHisPhe	95

US-08-477-451-3
Sequence 3, Application US/08477451
Patent No. 5928665
GENERAL INFORMATION:
APPLICANT: Corbett, Antonello
TITLE OF INVENTION: Helicobacter Pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3174 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-451-3

Alignment Scores:
Pred. No.: 3,116-87 Length: 3174
Score: 927.50 Matches: 215
Percent Similarity: 94.3% Conservative: 0
Best Local Similarity: 94.3% Mismatches: 0
Query Match: 79.2% Indels: 13
DB: 1 Gaps: 13

US-09-360-685C-26 (1-685) x US-08-477-451-3 (1-3174)

QY 2 AAAATGCGAAATAAGATTTCAGACGTAACGCAAGCAAAAAGCACTTGAATAAT 61
DB 2201 LysMetAlaLysIleArgIleSerAlaArg---ArgLysGlnLysAlaThrLysIle 2219
QY 62 CGTTAAAGATGATCATCATCAATCAAAAGGTAAACGTAAGTGAATATCTCAATCAG 121
DB 2220 ProLeuLysMet---SerSerIleLysArg---ArgIleLysLeuIleSerIleLys 2237
QY 122 CGGATACAGTGCCTAAAGCAACGGGTGATTTCAGTAGGAGTAAAGCAAGCGTTAGCCGATC 181
DB 2238 ArgTyrGlnTrpLysGlnArgValIleSerValGly---SerLysArg---ProIle 2255
QY 182 TCAAAAATTTCTCAAAAGCAAGCAATTCGCCCAACAGCTCAAAAATAAGTTCATG 241
DB 2256 SerLysIleSerGlnArgSerAsnTrpProAsnLysLeuLysLysMetLysValSerMet 2275
QY 242 CTAGAAAAAATCTGAAATATATCAATCGTTAAGAAAGTGTGAATGAACCTAGTCG 301
DB 2276 LeuGlnLysAsnLeuLysTyrIleAsnProLeuArgMetVal---MetGlnPro---Ser 2293
QY 302 GTAATGGGTTATCTCAAGCAAGCAACCACTCTTCTTAAAACTTTGACATCAAGA 361
DB 2294 ValMetGlyTyrLeuLysGlnLysProGlnLeuPheLeuLysThrPheArgThrSerArg 2313
QY 362 AAGAGTTGAATGCAAAACTTGAAATTCAATPAACAATAACATAATGACTCAAAAAGC 421
DB 2314 LysSer---MetGlnAsnLeuGlnIleSerIleThrIleMetAspSerLysThr 2332
QY 422 AACCAATTAAGTCTAAAGTATATAAAGAAGCAAGCGGCAAGCAAGTACCTTGAAGAC 481
DB 2333 AsnProPheMetLeuLysLeuLysArgLysGlnGlyLysGlnLeuAlaLeuLysAsn 2352
QY 482 CCATTACGCTCAAGTTGCTAAAAAAGTAAATGCAAAAATGACGCACTCAATCAAAATG 541
DB 2353 ProPheThrLeuLysLeuLysArg---MetGlnLysLeuThrAspSerIleLys--- 2370
QY 542 CAAGTGGTTGGGTGTGGAGGCAAGCGCGGCTTCCCTTTGAAAAGCAATGATTAAG 601
DB 2371 GlnValIleTrpValLeu---GlyLysGlnArgAlaSerLeu---LysGlnMetIleLys 2388
QY 602 TTGATGATCTCAGTAAGGTAGGCGTTTCAAGGAATCAAGAATTGCTCAAAAATTGACA 661

DB 2389 LeuMetIleSerValArg---GlyPheGlnGlyIleLysAsnTrpLeuArgLysLeuTr 2407
QY 662 ATCTCAATCAAGCGGTATCAAGAG 685
DB 2408 IleSerIleLysArgTyrGlnLys 2415

RESULT 10
US-08-053-614-4
Sequence 4, Application US/08053614
Patent No. 5403924
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
TITLE OF INVENTION: THE TGA GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/053,614
FILING DATE: 19930426
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-053-614-4

Alignment Scores:
Pred. No.: 9e-85 Length: 1181
Score: 902.50 Matches: 195
Percent Similarity: 76.1% Conservative: 6
Best Local Similarity: 73.9% Mismatches: 26
Query Match: 77.1% Indels: 37
DB: 1 Gaps: 4

US-09-360-685C-26 (1-685) x US-08-053-614-4 (1-1181)

QY 1 AAAATGCGAAATAAGATTTCAGACGTAACGCAAGCAAAAAGCACTTGAATAAT 60
DB 750 LysAsnGlyLysAsnLysAspPheSerLysValThrGlnAlaLysSerAspGlnLysAsn 769
QY 61 TCCGTTAAGAGTGCATCATCAATCAAAAGGTAAAGTGAATGTAATCTCAATCAA 120
DB 770 SerIleLysAspValIleIleAsnGlnLysIleThrAspLysValAspGlnLeuAsnGln 789
QY 121 GCGGTATCAGTGGCTAAAGCAACGCGGTGATTTCAGTAGGAGTAAAGCAAGCGTTAGCCGAT 180
DB 790 AlaValSerValAlaLysIleAlaCysAspPheSerGlyValGlnGlnAlaLeuAlaAsp 809
QY 181 CTCAAAAATTTCTCAAAAGCAAGCAATTCGCCCAACAGCTCAAAAATAAGTCTCAAT 240

Tue Mar 7 12:58:24 2006

us-09-360-685c-26.fai

Page 8

Db 810 LeuLyAsnPhseSerlysgLInleuAlaGlnInAlaGlnLyAsnGlnSerPheAsn 829
QY 241 GCTAGAAAAATCTGAATATATCATCGCTTAAGAGTGCGATGGAACCTGATCTC 300
Db 830 Val---GlyLySerGlnIleTyrgInSerValIleAsnGlnValAsnGlnIleVal 848
QY 301 GGTATGGGTATCTCAAGAGACCACTCTTCAAAAGCTTAAACCTTTGGCATCAAG 360
Db 849 GlyAsnGlyLeuSerGlyIleGlnAlaThrAlaLeuAlaLyAsnPhseSerAspIleAs 868
QY 361 AAAGGTTGAATGCAAACTTGGAATTTCAATTAACATTAATGACATCAAAAC 420
Db 869 LygGluLeuAsnGlnIlySpheLyAsnPhseAsnPhseAsnAsnAsnGlnLyLeuLyAsn 888
QY 421 -----GAACCATTTATGCTAAAGTTAATTAATTAAGAAAGAGGCAAGCATGACCTT 474
Db 889 GlyGlyIleProIleTyrgAlaGlnValAsnLyLeuLyThrGlyGlnValAlaSerPro 908
QY 475 GAAGAACCATTTACGCTCAAGTCTTAAAGGTAATGCAAAATTTGACGACTCAAT 534
Db 909 GlnGluProIleTyrgAlaGlnValAlaLyLeuValThrLySylleAspGlnLeuAsn 928
QY 535 CAATTAAGCA--AGTGGTTGGGTGTGTGAGGCAAGCA----- 570
Db 929 GlnAlaAlaThrSerGlyPheGlyValGlyGlnAlaGlyPheProLeuLyAsnGln 948
QY 570 ----- 570
Db 949 AspLyValGlnAspLeuSerLyValGlyArgSerValSerProGluProIleTyrgAla 968
QY 571 -----GCGGGCTTCCTCTTTGAAAAGGCAATGAATGATGATCTC 612
Db 969 ThrIleAspLeuGlyGlySerPheProLeuLyArgHIsAspLyValAspAspLeu 988
QY 613 AGTAAGTAGGGCTTTCAAGAGATCAAGAAATTTGGCTCAAGAAATTTGACATCTCAAA 672
Db 989 SerLyValGlyLeuSerArgAsnGlnIleuThrGlnIlySylleAspAsnLeuSerGln 1008
QY 673 GCGGTATCAGAA 684
Db 1009 AlaValSerGln 1012
RESULT 11
US-08-316-397B-4
; Sequence 4, Application US/08316397B
; Patent No. 5733740
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOS
; APPLICANT: TUMMURU, MURALI K. R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P. C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/316,397B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016

REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-316-397B-4
Alignment Scores:
Pred. No.: 9e-85 Length: 1181
Score: 902.50 Matches: 195
Percent Similarity: 76.1% Conservative: 6
Best Local Similarity: 73.9% Mismatches: 26
Query Match: 77.1% Indels: 37
DB: 1 Gaps: 4
US-09-360-685c-26 (1-685) x US-08-316-397B-4 (1-1181)
QY 1 AAAAATGGCAAAATTAAGATTTCAGCAAGTAAAGCAAGCAAAAGCACTTGAAT 60
Db 750 LyAsnGlyLyAsnLyAspPheSerLyValThrGlnAlaLySerPheGlnGln 769
QY 61 TCGTTAAAGATGTGATCATCATCAATCAAAAGCTAAGCAATTAATGATTAATCTCAATCA 120
Db 770 SerIleLyAspValIleIleAsnGlnIlyThrAspLyValAspGluLeuAsnGln 789
QY 121 GCGGTATACGCTTAAAGCAAGCGGATTTCAAGAGGTAGAGCAAGCTTACCGCAT 180
Db 790 AlaValSerValAlaLySylleAlaCyAspPheSerGlyValGlnGlnAlaLeuAlaAsp 809
QY 181 CTCAAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCAAAAATTAAGATCTCAAT 240
Db 810 LeuLyAsnPhseSerLysgLInleuAlaGlnInAlaGlnLyAsnGlnIleuSerPheAsn 829
QY 241 GCTAGAAAAATCTGAATATATCATCGCTTAAGAGTGCGATGGAACCTGATCTC 300
Db 830 Val---GlyLySerGlnIleTyrgInSerValIleAsnGlnValAsnGlnIleVal 848
QY 301 GGTATGGGTATCTCAAGAGACCACTCTTCAAAAGCTTAAACCTTTGGCATCAAG 360
Db 849 GlyAsnGlyLeuSerGlyIleGlnAlaThrAlaLeuAlaLyAsnPhseSerAspIleAs 868
QY 361 AAAGGTTGAATGCAAACTTGGAATTTCAATTAACATTAATGACATCAAAAC 420
Db 869 LygGluLeuAsnGlnIlySpheLyAsnPhseAsnPhseAsnAsnAsnGlnLyLeuLyAsn 888
QY 421 -----GAACCATTTATGCTAAAGTTAATTAATTAAGAAAGAGGCAAGCATGACCTT 474
Db 889 GlyGlyIleProIleTyrgAlaGlnValAsnLySylleThrGlyGlnValAlaSerPro 908
QY 475 GAAGAACCATTTACGCTCAAGTCTTAAAGGTAATGCAAAATTTGACGACTCAAT 534
Db 909 GlnGluProIleTyrgAlaGlnValAlaLyLeuValThrLySylleAspGlnLeuAsn 928
QY 535 CAATTAAGCA--AGTGGTTGGGTGTGTGAGGCAAGCA----- 570
Db 929 GlnAlaAlaThrSerGlyPheGlyGlyValGlyGlnAlaGlyPheProLeuLyAsnGln 948
QY 570 ----- 570
Db 949 AspLyValGlnAspLeuSerLyValGlyArgSerValSerProGluProIleTyrgAla 968
QY 571 -----GCGGGCTTCCTCTTTGAAAAGGCAATGAATGATGATCTC 612
Db 969 ThrIleAspLeuGlyGlySerPheProLeuLyAsnGlnIleuThrGlnIlySylleAspAspLeu 988
QY 613 AGTAAGTAGGGCTTTCAAGAGATCAAGAAATTTGGCTCAAGAAATTTGACATCTCAAA 672
Db 989 SerLyValGlyLeuSerArgAsnGlnIleuThrGlnIlySylleAspAsnLeuSerGln 1008

QY 673 GCGGTATCAGAA 684
Db 1009 AlavalSerGlu 1012

RESULT 12

US-09-034-306-4
Sequence 4, Application US/09034306
Patent No. 5876943
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: HARRY KLEANTHOS
APPLICANT: TUMMURU, MURALI K.R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,306
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200,030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-034-306-4

Alignment Scores:

Pred. No.: 9e-85 Length: 1181
Score: 902.50 Matches: 195
Percent Similarity: 76.1% Conservative: 6
Best Local Similarity: 77.9% Mismatches: 26
Query Match: 1 Indels: 37
Gaps: 4

US-09-360-685C-26 (1-685) x US-09-034-306-4 (1-1181)

QY 1 AAAAAATGCAAAATAGATTTCACGAGGTAACGACGCAAAAGCACTTGAAT 60
Db 750 LysAenGlyLysasnLysaspPheSerLysValThrGlnAlaLysSeraspGlnGluAan 769
QY 61 TCCGTTAAGATGTGATCATCATCAAAAGGTAAAGGATTAAGTTATATCATCA 120
Db 770 SerLleLysaspValIleIleasnGlnLysIleThrAspLysValaspGluLeuasnGln 789
QY 121 GCGGTATCAGTGGCTTAAGCAACGCGGTGATTTCAGTAGGATAGAGCAAGCGTTAGCCGAT 180
Db 790 AlaValSerValAlaLysIleAlaLysaspPheSerGlyValGlnGlnAlaLeuAlaasp 809

QY 181 CTCAAAAATTTCTCAAGAGCAATGGCCCAACAGTCCAAAAATGAAATCTCAAT 240
Db 810 LeuLysasnPheSerLysGlnLeuAlaGlnGlnAlaGlnLysasnGlnSerPheasn 829
QY 241 GCTAGAAAAAATCTGAAATATATCATCCGTTAAGAAATGCTGATGAAACCTAGTC 300
Db 830 Val---GlyLysSerIleIleThrGlnSerValLysasnGlyValasnGlyThrLeuVal 848
QY 301 GGTAAATGGTTATCTCAACGAGCAAGCCACACTCTTTCTTAAAACTTTTCGACATCAAG 360
Db 849 GlyasnGlyLeuSerGlyIleGlnAlaThrAlaLeuAlaLysasnPheSerPheIleLys 868
QY 361 AAGAGTTGAATGCCAACTTGGAATTTTCATTAACAATTAACAATTAAGCTCAAAAAC 420
Db 869 LysGlnLeuasnGlnLysPheLysasnPheasnAsnAsnAsnAsnGlnLysLysasn 888
QY 421 -----GAACCCATTATGCTAAAGTTAATTAATAAGCAAGGCAAGCAGCTACCTT 474
Db 889 GlyGlyGlnProIleThrAlaGlnValasnLysLysThrGlyGlnValAlaSerPro 908
QY 475 GAAGAACCCATTATGCTCAAGTTCGCTAAAGGTAATGCAAAATTAACGACTCAAT 534
Db 909 GlnGlnProIleThrAlaGlnValAlaLysValThrLysLysIleaspGlnLeuasn 928
QY 535 CAATATGCA---AGTGGTTGGGTGTTAGGGCAAGCA----- 570
Db 929 GlnAlaAlaThrSerGlyPheGlyValGlyGlnAlaGlyPheProLeuLysArgHis 948
QY 570 ----- 570
Db 949 AspLysValGlnAspLeuSerLysValGlyArgSerValSerProGlnProIleThrAla 968
QY 571 -----GCGGCTTCCTTTGAAAGGATGATTAAGTTGATGATCTC 612
Db 969 ThrIleAspAspLeuGlyLysSerPheProLeuLysArgHisAspLysValaspLeu 988
QY 613 AGTAAGTGGGCTTTCAGAGGAATCAAGAAATGGCTCGAAATTAATGACATCTCAATCA 672
Db 989 SerLysValGlyLeuSerArgasnGlnGlnLeuThrGlnLysIleaspAsnLeuSerGln 1008
QY 673 GCGGTATCAGAA 684
Db 1009 AlavalSerGlu 1012

RESULT 13

US-09-259-437-4
Sequence 4, Application US/09259437
Patent No. 6153390

GENERAL INFORMATION:

APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: HARRY KLEANTHOS
APPLICANT: TUMMURU, MURALI K.R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/259,437
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US/08/316,397
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GRENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-9880
TELEFAX: 404/688-0770
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-259-437-4

Alignment Scores:
Pred. No.: 9e-85 Length: 1181
Score: 902.50 Matches: 195
Percent Similarity: 76.1% Conservative: 6
Best Local Similarity: 73.9% Mismatches: 26
Query Match: 77.1% Indels: 37
DB: 2 Gaps: 4

US-09-360-685C-26 (1-685) x US-09-259-437-4 (1-1181)

```
QY 1 AAAAATGGCAAAATTAAGATTTCAGCAAGTAAACGCAAAAGCGACTTGAAT 60
DB 750 LysAenGlyLysAenLysAenPheSerLysValThrGlnAlaLysSerAenGlnLysAen 769
QY 61 TCCGTTAAAGATGTCATCATCAATCAAAAGGTAAACGATTAAGTTGATATCTCAATCA 120
DB 770 SerLysAenPheValLLeileAenGlnLysLLeThrAenPheValAenPheGlnLysAen 789
QY 121 GCGGTATCAGTGGCTTAAGCAACGCGTATTCAGTAAAGGTAGAGCAAGCGTTAGCCGAT 180
DB 790 AlaValSerValAlaLysLLeAlaCysAenPheSerGlyValGlnGlnAlaLysAen 809
QY 181 CTCGTTAAATTTCTCAAAAGCAAGATTCGCTCAACCAAGCTCAAAATGAAGTCTCAAT 240
DB 810 LysLysAenPheSerLysGlnLysAenLysAenLysAenLysAenLysAenLysAen 829
QY 241 GCTAGAAAAAATCTGAATATATCAATCCGTTAAGATGCTGAAGTGAACCTGCTGTC 300
DB 830 Val---GlyLysSerGlnLysLLeGlnSerValLysAenGlyLysAenGlyLysAen 848
QY 301 GGTAAATGGTATCTCAAGCAAGCAAGCAAGCTCTTCTAATAAACTTTGCGACATCAAG 360
DB 849 GlyAenGlyLysSerGlyLysLLeAlaThrAlaLysAenLysAenPheSerLysAen 868
QY 361 AAAGAGTTGAATGCAGAACTTGAAATTTCAATTAACATTAACATTAATGACTCAAAAC 420
DB 869 LysGlnLysAenGlnLysPheLysAenPheAenAenAenAenAenAenAenLysAen 888
QY 421 -----GAACCATTTAGTAAAGTTAAGTTAATAAAAGCAAGCGCAAGCTAGCTTC 474
DB 889 GlyGlyLysPheLysLLeGlnLysAenLysAenLysAenLysAenLysAenLysAen 908
QY 475 GAAGAACCATTTAGCTCAAGTCTCAATTAAGTTAATAAGCAAAATTTGACCGACTCAAT 534
DB 909 GlnGlnPheLysLLeGlnLysAenLysAenLysAenLysAenLysAenLysAenLysAen 928
QY 535 CAATTAAGCA---AGTCGTTGGGTTGGGTTAGGCGCAAGCA--- 570
DB 929 GlnAlaAlaThrSerGlyPheGlyGlyValGlyGlnAlaGlyPheProLysAenGlnHis 948
QY 570 ----- 570
DB 949 AspLysValGlnAenPheSerLysValGlyAenSerValSerProGlnPheLysAla 968
QY 571 -----GCGGCGTTCCCTTTGAAAAGGCAATTAAGTTAGTGAATCTTC 612
```

```
DB 969 ThrLeaAenPheLysGlyLysPheProLysAenGlnHisAenPheValAenAenPheLys 988
QY 613 AAGTAGTAGGCTTTCAAGCAATTCAGCAAGTTCGCTCAAGAAATTTGCAATCTCAATCA 672
DB 989 SerLysValGlyLysSerAenGlnLysAenGlnLysAenGlnLysAenLysAenLysAen 1008
QY 673 GCGGTATCAGCA 684
DB 1009 AlaValSerGln 1012

RESULT 14
PCT-US93-09782-4
Sequence 4, Application PC/TUS9309782
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/09782
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GRENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-9880
TELEFAX: 404/688-0770
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-09782-4

Alignment Scores:
Pred. No.: 9e-85 Length: 1181
Score: 902.50 Matches: 195
Percent Similarity: 76.1% Conservative: 6
Best Local Similarity: 73.9% Mismatches: 26
Query Match: 77.1% Indels: 37
DB: 4 Gaps: 4

US-09-360-685C-26 (1-685) x PCT-US93-09782-4 (1-1181)
QY 1 AAAAATGGCAAAATTAAGATTTCAGCAAGTAAACGCAAAAGCGACTTGAAT 60
DB 750 LysAenGlyLysAenLysAenPheSerLysValThrGlnAlaLysSerAenGlnLysAen 769
QY 61 TCCGTTAAAGATGTCATCATCAATCAAAAGGTAAACGATTAAGTTGATATCTCAATCA 120
DB 770 SerLysAenPheValLLeileAenGlnLysLLeThrAenPheValAenPheGlnLysAen 789
QY 121 GCGGTATCAGTGGCTTAAGCAACGCGTATTCAGTAAAGGTAGAGCAAGCGTTAGCCGAT 180
DB 790 AlaValSerValAlaLysLLeAlaCysAenPheSerGlyValGlnGlnAlaLysAen 809
```

RESULT 15
 US-08-477-451-4
 ? Sequence 4, Application US/08477451
 ? Patent No. 5928865
 ? GENERAL INFORMATION:
 ? APPLICANT: Covacci, Antonello
 ? TITLE OF INVENTION: Helicobacter Pylori CagI Region
 ? NUMBER OF SEQUENCES: 46
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Chiron Corporation
 ? STREET: 4560 Horton Street
 ? CITY: Emeryville
 ? STATE: CA
 ? COUNTRY: USA
 ? ZIP: 94608-2916
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Floppy disk
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: PatentIn Release #1.0, Version #1.30
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/08/477,451
 ? FILING DATE: 07-JUN-1995
 ? CLASSIFICATION: 435
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: McClung, Barbara G.
 ? REGISTRATION NUMBER: 33,113
 ? REFERENCE/DOCKET NUMBER: 0335.002
 ? TELECOMMUNICATION INFORMATION:

QY	3	AAATGGCAAAATAAGATTTCAGCAAGATGAACCAAGCAAAAGACACTTGAAAATTC	62
Db	2217	LysTrpGlnLys---GlyPheGlnGlnCylAsnHisLeuSerLysLysArgPro--LysPhe	2234
QY	63	CGTTAAAGATGTGATCATCATCAAAAGTGTAACGGATAAAGTTGATTAATCATCAATCAAC	122
Db	2235	Arg---ArgCysAspHisLgInSerLysGlyAsnIly-----SerSerGlnSerSer	2250
QY	123	GGTATCGATGGCTTAACCAACGGGTGATTTCAAGTGGGTAGACCAACGTTAGCCGATCT	182
Db	2251	GlyIleSerGly---SerAsnGly--PheGln--GlyArgAlaSerValSerArgSer	2267
QY	183	CAAAAATTTCTCAAGGAGCAATTTGGCCCAACAAGCTCAAAAAATGAAGTCTCAATGC	242
Db	2268	GlnLysPheLeuLysGlyAlaIleGlyProThrSerSerLysLys--LysSerGlnCys	2286
QY	243	TAGAAAAAATCTGAATTAATATCAATCCGTTAAGATGGTGTAAATGGAAACCTCATGTCG	302
Db	2287	---LysLysIle---AsnIleSerIleArg--GluTrpCysGluTrpAsnProSerArg	2303
QY	303	TAAATGGTATCTCAACGACAGCAACCACTCTTCTTAAAAACTTTCCGACATCAAGAA	362
Db	2304	---TrpValIleSerSerArgSerHisAsnSerPhe--LysLeuPheGlyHisGlnGlu	2321
QY	363	AGAGTTAAATGCAAAACTTGGAAATTTCAATTAACAATAACAATATGACTCAAAAACGA	422
Db	2322	ArgValGlnCysLysTrpLysPheGln-----GlnGlnTrpThrGlnLysArg	2338
QY	423	ACCAATTATGCTTAAAGTTAATAAAGAAAGCAGGGCAAGCAAGCTAGCCTGGAAGACC	482
Db	2339	ThrHisLeuLys-----SerLysGlnSerArgAlaSerSer--Pro--ArgTrp	2353
QY	483	CATTACGCTCAAGTTCCTAAAAAGATAAATGCAAAAATTTGACCGACTCAATCAAAATGC	542
Db	2354	HisLeuArgSerSerCys--LysGlyLysCysLeuAsn--ProThrGlnSerAsnSer	2371
QY	543	AAGGTGTTGGGTGTGAGGCAAGAGAGGGGCTCCCTTGTAAAAAGCATATAAGT	602
Db	2372	LysTrpPheLysCysArgAlaSerSerGlyLeuProPheGlnLysAla-----	2388
QY	603	TGATGATCTCAAGTAAAGGTAGGGCTTTCAAGAAATCAAGATTTGGCTCAGAAAATTCGAA	662
Db	2389	-----SerSerGlnGlyArgAlaPheLysGlnSerArgIleGlySerGluAsn--Gln	2405
QY	663	TCTCAATCAAGCGGTATCAAG	683
Db	2406	SerGlnSerSerGlyIleArg	2412

This Page Blank (uspto)

XX

PN WO318150-A1.
 XX 16-SEP-1993.
 PD 02-MAR-1993; 93WO-EP000472.
 XX 02-MAR-1993; 92IT-F1000052.
 PR 25-JAN-1993; 93WO-EP000158.
 XX (ISTS) BIOGINE SCLAVO SPA.
 PA
 PI Covacci A, Bugnoli M, Telford J, Macchia G, Rappuoli R;
 XX WPI; 1993-303464/38.
 DR N-PSDB; AAQ48733.
 PT Recombinant Helicobacter pylori protein and corresp. gene - is a
 PT cytotoxin, antigen or heat shock protein used for treating and preventing
 PT type B gastritis, gastric ulcers and gastric tumours.
 PS Claim 5; Fig 4; 83pp; English.
 CC This sequence represents the cytotoxin-associated ammunodominant antigen,
 CC CAI, of H. pylori. The protein is a hydrophilic, surface-exposed protein,
 CC having a molecular weight of approximately 120-132 kD. Ref. 128-130 kD,
 CC and an isoelectric point of 9.72. H. pylori isolates which do not
 CC produce the CAI antigen do not have the cai gene, and are also unable to
 CC produce an active cytotoxin. The association between the presence of the
 CC cai gene and cytotoxicity suggests that the product of the cai gene is
 CC necessary for the translocation, folding, export or function of the
 CC cytotoxin. Alternatively both the cai gene and the ct gene are absent in
 CC noncytotoxic strains suggesting physical linkage between the genes. The
 CC absence of a typical leader sequence suggests the presence of an
 CC independent export system. Computer searches for promoter regions in the
 CC region upstream from the ATG codon identified sequences resembling either
 CC -10 or -35 regions, however a region with a good consensus to the E. coli
 CC H. pylori promoter sequences was not found. The CAI antigen is very
 CC hydrophilic and does not show obvious leader peptide or transmembrane
 CC sequences. The most hydrophilic region is from amino acids 600-900, where
 CC a number of unusual features are observed. There is a repetition of the
 CC contiguous asparagines. This protein, and the presence of a stretch of six
 CC esp. cytotoxin (CT) or a heat shock protein (see also AAR41198 and
 CC AAR41200), may be used to treat, prevent and diagnose H. pylori
 CC infection. H. pylori is the causative agent of type B gastritis, peptic
 CC ulcers and gastric tumours. (Updated on 25-MAR-2003 to correct PN field.)
 CC (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to
 CC correct PA field.)
 CC
 SQ Sequence 1147 AA;
 Alignment Scores:
 Pred. No.: 7.15e-106 Length: 1147
 Score: 1127.00 Matches: 228
 Percent Similarity: 100.0% Conserved: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 96.2% Indels: 0
 DB: 2 Gaps: 0
 US-09-360-685c-26 (1-685) x AAR41199 (1-1147)
 QY 1 AAAAAAGCGAAAAATGAGATTTCAGCAAGTAACGCAAGCAAAAAAGCACTTGAAT 60
 Db 750 LysaengilYlYahshuYahpneSerlyValthrlhAlYseSerAepLengluNen 769
 QY 61 TCGGTTAAGATGATCATCATCAATCAAAAGTAAACGATTAAGTGAATATTCATCA 120
 Db 770 SerVailYahpValIleleahengInlyValthhAspLyValAaPhaenLenuIn 789
 QY 121 GCGGATCAGTGGCTTAAGCAACGCGTGAATTCAGTAGGCTAGCAAGCAAGTTCAGCAT 180
 Db 790 AlavaiSerValahAlahValahthrlYahpneSerAryValGIugInlaleNulap 809

QY 181 CTCAAAAATTTCTCAAGAGCAATTGGCCCAACCTCAAAAAATGAATCTCAAT 240
 Db 810 LeuYahpneSerlyGluInlenuIleGInlYahnglYseSerLeuNen 829
 QY 241 GCTAGAAAAAATCTGAATATATCAATTCGCTTAAGATGGTGAATGAACCTTAGTC 300
 Db 830 AlahryGlyYseSerGluIleYrGInSerValYahnglYValahnglYthrlYVal 849
 QY 301 GGTAAATGGGTATCTCAAGAGCAACCACTTTCTCAAAAACTTTGGCAATCAAG 360
 Db 850 GlyYahnglYleuSerGInlaleGInlathrlhLeuSerlyAahpneSerApyllyls 869
 QY 361 AAAGATTGAATGCAAACTTGCAAAATTTCAATACATTAACATTAATGACTCAAAAAC 420
 Db 870 LyGluLeuAahAlahYleuGInlYahpneAahAahAahAahAahnglYleuYahN 889
 QY 421 GAACCAATTTATGCTAATTAATTAATTAAGCAAGCGCAAGCACTAGCTTGAAGAA 480
 Db 890 GluProIleYrAlahYValahNulYahYValahnglYlnAlahSerLeuGlu 909
 QY 481 CCGATTAGGCTCAAGTCTTCAAAAGCTTAATGAATTAACCACTCAATCAATTA 540
 Db 910 ProIleYrAlahngInlAlahYValahNulYahNulYahNulYahNulYahNul 929
 QY 541 GCAAGTGGTTGGCTGTTGAGCAAGCGGCGCTTCCCTTGAAGGATGATTA 600
 Db 930 AlahSerGlyleuGlyValahGlyGlnAlahAlahnglYpneProleuYahRhlAaply 949
 QY 601 GTTGATATCTCAAGTGAAGTGGCTTCAAGCAATCAAGATTCGATCAAAATTCAC 660
 Db 950 ValahpneSerlyValahGlyleuSerAryahngInlYahnglYleuYahp 969
 QY 661 AATCTCAATCAACCGCTTCAAGAA 684
 Db 970 AahleuAahngInlAlahSerGln 977
 RESULT 2
 ID AAM80600 standard; protein; 1338 AA.
 AC AAM80600;
 AC 17-OCT-2003 (revised)
 DT 15-FEB-1999 (first entry)
 XX
 DE Helicobacter pylori Caga-Vibrio cholerae A2+B subunit fusion.
 XX
 KW Vaccine; antigen; Caga; antigen; toxin; diagnosis; gastritis; ulcer;
 KW stomach cancer.
 OS Helicobacter pylori.
 OS Vibrio cholerae.
 OS Chimeric.
 FH
 FT Key
 FT Protein
 FT 1. 1147
 FT 1148. 1338
 FT /label= Caga
 FT /label= A2+B_subunits
 FN
 PN WO9844130-A1.
 PD 08-OCT-1998.
 XX 31-MAR-1998; 98WO-KR000073.
 XX 31-MAR-1997; 97KR-00011950.
 PR 31-MAR-1997; 97KR-00011951.
 XX (DAEW-) DAEWOONG PHARM CO LTD.
 PA
 PI Kim B, Shin S, Yu Y, Park M, Choi D, Jung H,

DR WPI; 1998-568279/48
DR N-PSDB; AAV62461.

DR N-PSDB; AAV62461

PT New chimeric proteins for use against *Helicobacter pylori* - comprising an
PT antigenic protein of *H. pylori* and A1 and B subunits of *Vibrio cholerae*
PT toxin, preferably produced by recombinant techniques.
XX

PT	toxin, preferably
XX	

PS Claim 8; Page 83; 102pp; English.
vvy

The invention relates to: chimeric proteins comprising antigenic proteins of *H. pylori* and A2 and B subunits of *V. cholerae* toxin; recombinant DNAs encoding such chimeric proteins; recombinant expression vectors; a process for preparing the chimeric proteins using transformed microbial host cells; and to preventative and therapeutic vaccines comprising the chimeric proteins for *H. pylori*-associated diseases such as gastritis, gastric ulcer, duodenal ulcer and gastric cancer (claimed). The chimeric proteins are designed to have excellent immunogenicity, to be stable in the stomach, to penetrate the mucous membrane of the intestine, and to stimulate production of sIgA. They can additionally be used as active ingredients in diagnostic kits for *H. pylori* infection, and for production of anti-*H. pylori* antibody. (Updated on 17-OCT-2003 to standardise OS field)

.SQ Sequence 1338 AA;

Alignment Scores:

Pred. No.:	7.44e-106
Score:	1127.00

```
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
```

Query Match: 96.2%

DB: 2

US-09-360-685C-26 (1-685) x AA#80

QY 1 AAAAATGCGAAATAAGG

Db 750 LYSASNGLYLYSASNLysA

61 TCCGTTAAGATGTGATCA

Db 770 ServaIv8ASpVallier

121 සලසාකළ

[illegible]

790 AlavalSerValAlalysa

US-09-360-685C-26 (1-685) x AAW80600 (1-1338)

QY	AAAAATGGCAAAATATAGATTTCAGCAAGGTAAAGCAAGCAAAAGGACCTTGAAAT	60
Db	LYSANGLYLysaSnLYsaPhSeSerLYsValThrGlnAlaLYSeSerAspLeuGlnsn	76
QY	61 TCCGTTAAAGATCTGATTCATCAATCAAAAGGTAAAGGATTAAGGTGATATCTCAATCA	120
Db	770 SerValLYsaSPaIllelleaSnGlnLYsValThrAspLYsaAspAsnLeuasnGln	785
QY	121 GCGGTATCAGTGCGCTTAAGCAACGGGTGATTTTCAGTAGGGTGTAGCAACGGTTAGCGCAT	180
Db	790 AlaValSerValAlaLYsaIaThrGlyAspPhSeSerArgValGluGlnAlaLeuAlaAsp	805
QY	181 CTCAAAAATTTCTCAAGAGAGCAATTGGCCCAACAAGCTCAAAAAATGAAGTCTCAT	240
Db	810 LeuLYsaSnPhSeSerLYsGluGlnLeuAlaGlnGlnAlaGlnLYsaSnGlySerLeuasn	825
QY	241 GCTAGAAAAAAATCTGAATATATATCAATCCGTTTAAAGATGGGTAAAGAAACCTTAGTC	300
Db	830 AlaArgLYsLYSeSerCulIeTYrGlnSerValLYsaSnGlyValaSnGlyThrLeuVal	845
QY	301 GGTAAATGGTTATCTCAGACGAGAGCCACAACCTTTTCTTAAAACTTTTGCAATCAG	360
Db	850 GlyaSnGlyLeuSerGlnIaGlnAlaThrThrLeuSerLYsaSnPhSeSerAspIleLYs	865
QY	361 AAAGATTTGAATGCANAAACTTGGAAATTCATTAACATTAACAAATAAGACTCAAAAAC	420
Db	870 LYsGluLeuAsnIaLYsLYSeuGlyAsnPhleAsnAsnAsnAsnAsnAsnGlyLeuLYsaSn	885
QY	421 GAAACCATTTATGCTTAAAGTTAATAAAAAAGAAAGGAGGCAACGACTTACCTTGAAGAA	480
Db	890 GluProIleTYrAlaLYsaValaSnLYsLYsLYsValaGlyGlnAlaIaSerLeuGln	905
QY	481 CCCATTAGGCTCAAGTTGCTAAAAAGGTAAATGGAATAATGACCCGACTCATCAATA	540
Db	910 ProIleTYrAlaGlnValaLYsLYsValaSnAlaLYsIleAspArgLeuAsnGlnIle	925

Qy	541	GCAGAGGGTTTGGGTTTGTAGGGCAAGAGGGGCTTCCCTTTGAAGAAGCATGTAA	600
Dd	930	AlaserGlyLeuGlyValValGlyGlnAlaAlaGlyPheProLeuLysArgHisAspLys	949
Qy	601	GTTCATGATCTCAGTAGGTAGGGCTTTTCAGAGATCAAGAAATTTGCTCAGAAATTTGAC	660
Dd	950	ValAspAspLeuSerLysValGlyLeuSerArgAsnGlnGluLeuAlaGlnLysIleAsp	965
Qy	661	AATCTCAATCAAGCGGTATCAGAA	684
Dd	970	AsnLeuAsnGlnAlaValSerGlu	977
RESULT 3			
ID	ADU05584	standard; protein; 1186 AA.	
XX	ADU05584;		
AC	ADU05584;		
XX	27-JAN-2005	(first entry)	
DT	XX		
XX	H. pylori	antigenic protein HP0547.	
DE	XX		
XX	antibacterial; antigenic; H. pylori	infection.	
OS	XX	Helicobacter pylori.	
XX	Key	Location/Qualifiers	
FT	Region	110..357	
FT	FT	/note= "Immunogenic region. Region specifically claimed	
FT	FT	in claim 14"	
FT	Region	358..501	
FT	FT	/note= "Immunogenic region. Region specifically claimed	
FT	FT	in claim 14"	
FT	Region	502..1161	
FT	FT	/note= "Immunogenic region. Region specifically claimed	
FT	FT	in claim 14"	
PN	XX	MO2004094467-A2.	
XX	XX		
PD	XX	04-NOV-2004.	
XX	XX		
PF	XX	22-APR-2004; 2004MO-EP004255.	
XX	XX		
PR	XX	22-APR-2003; 2003EP-00450097.	
XX	XX		
PA	XX	(INTE-) INTERCELL AG.	
PI	XX		
PI	XX	Weinke A, Min Bui D, Nagy E, Henics T;	
DR	XX	WPI; 2004-775908/76.	
DR	XX	N-PSDB; ADU05406.	
XX	XX		
PT	XX	New hyperimmune serum reactive antigens from Helicobacter pylori, and	
PT	XX	encoding nucleic acid molecules, useful for diagnosing, preventing or	
PT	XX	treating H. pylori infections.	
PS	XX	Claim 14; SEQ ID NO 232; 176bp; English.	
XX	XX		
CC	CC	The invention relates to an isolated nucleic acid molecule encoding a	
CC	CC	hyperimmune serum reactive antigen or its fragment. The composition	
CC	CC	(including the nucleic acid molecule, hyperimmune serum-reactive antigen	
CC	CC	or antibody) is useful for manufacturing a medicament or pharmaceutical	
CC	CC	preparation (e.g. a vaccine) for treating or preventing H. pylori	
CC	CC	infections. The antigen or its fragment may also be used for isolating,	
CC	CC	purifying and/or identifying an interaction partner of the hyperimmune	
CC	CC	serum reactive antigen or fragment; for generating a peptide binding to	
CC	CC	the hyperimmune serum reactive antigen or fragment, where the peptide is	
CC	CC	selected from aptamers and spiegelmers; or for manufacturing a functional	
CC	CC	nucleic acid selected from ribozymes, antisense nucleic acids and	
CC	CC	siRNA. The present sequence represents the amino acid sequence of an H.	
CC	CC	pylori antigenic protein.	

CC	ulceration. A ligand e.g. Antibody, specifically reactive with the tag A
CC	antigen can be used to treat peptic ulcers. (Updated on 25-MAR-2003 to
CC	correct PN field.)
XX	
SQL	Sequence 1181 AA;
Alignment Scores:	
Pred. No.:	5.98e-83
Score:	902.50
Percent Similarity:	76.1%
Best Local Similarity:	73.9%
Query Match:	77.1%
DB:	2
	Gaps: 4
US-09-360-685C-26 (1-685) x AAR53269 (1-1181)	
Qy	1 AAAAAAGCAAAATAAGATTTCAGACGAGTAAAGCAAGCAAAAGCGACTTGAAT 60
Db	750 Lysanngilybaanlysaeppheserilyaithrginalalyseraepolnubsn 769
Qy	61 TCCGTTAAAGATGTGATCATCATCAATCAAAAGGTAAAGCATTAAGTTGATTAATCTCAATCAA 120
Db	770 Serilleysaepvalillelleasnghlylethrapslyvalaepgluleuasnghn 789
Qy	121 GCGGTATCAGTGCGTTAAAGCAACGGGTGATTTCAGTAAAGGTAGAGCAAGCGTTAGCCGAT 180
Db	790 Alalavalsevalialalyillealicyasapheaserilyalglunginalaleuhalasap 809
Qy	181 CTCAAAATTTCTCAAAAGAGCATTTGGCCCAACAGCTCAAAAAAATGAAGTCTCAAT 240
Db	810 Leulysanpheserlysglunglneulacnglnalaglnlysaenqluserpheasn 829
Qy	241 GCTAAGAAAAAATCTGAATAATATCAATCCGTTAAGATGGTGTGAATGGAACCTTACTC 300

[illegible]

Tue Mar 7 12:58:24 2006

us-09-360-685c-26.rag

Page 6

AA072593	AA072593 standard; protein, 1181 AA.
AC	AA072593;
AD	
AE	
AF	25-MAR-2003 (revised)
AG	29-SEP-1995 (first entry)
AH	
AI	H. pylori taga antigen.
AJ	
AK	Taga; antigen; ulcer; diagnosis; vaccine.
AL	
AM	Helicobacter pylori.
AN	US5403924-A.
AO	
AP	04-APR-1995.
AQ	
AR	26-APR-1993; 93US-00053614.
AS	
AT	13-OCT-1992; 92US-00959940.
AV	(UYVA-) UNIV VANDERBILT.
AW	
AX	Blaser MJ, Tummuru MKR, Cover TL;
AY	WPI; 1995-146855/19.
AZ	N-PSDB; AA086728.
BA	
BB	
BC	
BD	
BE	
BF	
BG	
BH	
BI	
BJ	
BK	
BL	
BM	
BN	
BO	
BP	
BQ	
BR	
BS	
BT	
BV	
BW	
BX	
BY	
BZ	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	

QY	AAAAATGGCAAAAATATAGATTTCACAGATTAACGACGACAAAACCGACCTGTAAT	60
Db	LYSAENGLYSAENGLYSAENPHESETRYVALTHGINALALYSERAPDGLIATAEN	76
QY	61 TCCGTTAAGATGATGATCATCATCAAAAGTAAACGATTAAGTTGATATCTCAATCA	120
Db	770 SerIIEYSAEPVALIETLEENGLINLYSLETHASPLVALSPDILUEANGLIN	781
QY	121 GCGGATTCATGCGCTAAAGACAGCGGTATTTTCATGAGGTAGACCAAGCTTACCGAT	180
Db	790 ALALVALSERVALALALYLLIETALCYAPHESETGVALGLUNGINLALUEALAAP	801
QY	181 CTCAAAAATTTCTCAAGAGACGAATGGCCCAACAGCTCAAAAAATGAAAGTCTCAAT	240
Db	810 LEUYSAENPHESETRYSGIUNGINLEUALGLINGINLALYSAENGLISEPHEAN	825
QY	241 GCTGAAAAAAAATCTGAAATAATCAATCCGTTAAGATGTTGTAATGACCAACCTGATC	300
Db	830 VAL---GLYVSAEGLIETRYGLINSEVALYSAENGLYALASNGLYTHLEUAL	846
QY	301 GGTAAATGGTTATCTCAAGACAGAACCAACCTTTTCTAAAACTTTGGACATCAAG	360

Dd 849 GlyncdiVleusSergIyIleGlnAlaThraLeuAlaLysPheSerPheIleIlys 866

OY 361 AAGAGTTGAATGCMAAACTTGAAATTTCATAACATTAACAATATGACTCAAAAC 422

Dd 869 LyscltLeuaenGluLySPhelySaenPheaenbenbnshnshnglyLeuLySan 886

OY 421 ----- GAACCATTATCTTAAAGTAAATTAATAAAAAAAGAACGGCAGCACGTACCTT 474

Dd 889 GlYglYGlupProiletryrInlaeGlnAlaMetuLySyryShrdIyGlnValaIsarPer 908

OY 475 GAAGAACCATTTAGCGCTCAAGTTCCTAAAAGSTAAATGCMAAAATTGACCGACATCAT 534

Dd 909 GlUGlupProiletryrInlaeGlnAlaLysValyValThrIleYalyleapGlnLeuen 928

OY 535 CAATATGCA---AGTGGTTGGGTGTGTGAGGCCACAA----- 570

Dd 929 GlnAlaIathrseryIphneGIyGIyGlnAlaGLIyPheProleuLyArghis 948

OY 570 ----- 570

Dd 949 AspLySVaIGlnAspleuSerLyVaIGlyArserValserProGluPeoIletryrInla 968

OY 571 ----- GGCGGCTCCCTTTGAAAAGCATGATTAAGTTGATGATCTC 612

Dd 969 ThrIleAspLeuGluGIyLyserePheProleuLyArghIsaspleySVaIAspAspleu 988

OY 613 AGTAAGAGGCGCTTCAAGAAATCAAGATGCGCTCAGAAATTTGACATCTCATCAAC 672

Dd 989 SerLySVaIGlyLeusehrgangelnGlnleuthGlnLySIleAspAmLeusenGln 1008

OY 673 GCGGATACCAA 684

Dd 1009 AlaValsercdu 1012

XX	AA91307;
XX	09-SEP-1996 (first entry)
XX	Helicobacter pylori TagA 120-128 kD antigen.
XX	Antigen; peptic ulcer; chronic gastritis; gastric adenocarcinoma;
XX	diagnosis; predileposition; antibody; vaccine; infection.
XX	Helicobacter pylori.
XX	WO9610639-A2.
XX	11-APR-1996.
XX	29-SEP-1995; 95WO-US012669.
XX	30-SEP-1994; 94US-00316397.
XX	(UYVA-) UNIV. VANDERBILT.
XX	(ORAV-) ORAVAX INC.
XX	Cover TL, Blaser MJ, Kleantous H, Tummuru MKR;
XX	WPI; 1996-209361/21.
XX	N-PSDB; AAT14051.
XX	Helicobacter pylori Tag A gene - used to develop and gastric
XX	carcinoma.
XX	Claim 4; Page 103-107; 118pp; English.
XX	AA91307 is a 120-128 kD antigen of H. pylori.

CC is designated TagA and its, or its fragments, can be used to determine a
 CC predilection to peptic ulceration or gastric carcinoma, conditions
 CC caused by or linked to HP infection. TagA and its fragments may also be
 CC used for antibody prodn. for use in detection of TagA in patients
 CC suspected of HP infection. TagA antibodies or other ligands may also be
 CC used to treat peptic ulceration or gastric carcinoma caused by HP
 CC infection. TagA or a non-functional TagA mutant may be used in vaccines
 CC for preventing and treating HP infection

XX
 XX Sequence 1181 AA:

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	5,98e-83	1181	195	6	26	37
Percent Similarity:	902.50					
Best Local Similarity:	76.1%					
Query Match:	73.9%					
DB:	77.1%					

US-09-360-685C-26 (1-685) x AAR91307 (1-1181)

QY 1 AAAAATGGCAAAATAGATTTTCAGCAAGTACCGCAAGCAAAAGCGACTTGAAT 60
 DB 750 LysAsnGlyLysAsnLysAspPheSerLysValThrGlnAlaLysSerAspGlnLysVal 769
 QY 61 TCCGTTAAAGATGATCATCAATCAAAAGGTAACGGTAAGTGAATATCTCAATCA 120
 DB 770 SerIleLysAspValIleIleAsnGlnLysIleThrAspLysValAspGlnLysVal 789
 QY 121 GCGGATCAAGTGGCTAAAGCAAGCGGTGATTTCAAGTAGAGCAAGCGGTAGCGAT 180
 DB 790 AlaValSerValAlaLysIleAlaCysAspPheSerGlyValGlnGlnAlaLysVal 809
 QY 181 CTCGAAATTTCTCAAGCAAGCAATGGCCCAACAAAGCTCAAAAATGAAAGTCTCAAT 240
 DB 810 LeuLysAsnPheSerLysGlnLysValGlnGlnAlaGlnLysAsnGlnSerPheAsn 829
 QY 241 GCTAGAAAAAATCTGAATATATCAATCCGTTAAAGATGCTGATGAAACCTGATC 300
 DB 830 Val--GlyLysSerLysIleLysValGlnSerValLysAsnGlyValAsnGlyThrLeuVal 848
 QY 301 GGTATGGGTTATCTCAAGCAAGCAAGCAAGCTTTCTTCAAAAATTTTCGACATCAAG 360
 DB 849 GlyAsnGlyLysSerLysIleGlnLysValIleAlaLysValAsnPheSerPheLys 868
 QY 361 AAAAGTTGAATGCAAACTGGAAATTTCAATACATTAACATTAATGACTCAAAAAC 420
 DB 869 LysGlnLysAsnGlnLysPheLysAsnPheAsnAsnAsnAsnAsnGlnLysValAsn 888
 QY 421 -----GACCATTATATCTAAAGTAAATTAAGCAAGCAAGCAAGCAAGCTT 474
 DB 888 GlyLysGlnLysIleLysValGlnLysValLysValLysValLysValLysValLysVal 908
 QY 475 GAAGAACCCATTACGCTCAAGTGTCTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 534
 DB 909 GlnGlnLysIleLysValGlnLysValLysValLysValLysValLysValLysValLysVal 928
 QY 535 CAATATGCA--AGTGGTTGGCTGTTGAGGCAAGCA----- 570
 DB 929 GlnAlaLysThrSerLysPheGlyValGlnLysValGlnLysValGlnLysValGlnLysVal 948
 QY 570 ----- 570
 DB 949 AspLysValGlnAspLysSerLysValGlnLysValGlnLysValGlnLysValGlnLysVal 968
 QY 571 -----GCGGGCTTCCTTTGAAAGGCAATGATTAAGTGAATGATTC 612
 DB 969 ThrIleAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 988
 QY 613 AGTAAGTGGGCTTCAAGCAAGCAAGTGGCTCAAGAAATGCAATCTCAATCA 672
 DB 989 SerLysValGlnLysSerLysAsnGlnLysLysLysLysLysLysLysLysLysLysLysLys 1008

QY 673 GCGGTATCAGAA 684
 DB 1009 AlaValSerGln 1012

RESULT 8

ID ABU51367 standard; protein; 420 AA.

AC ABU51367;

DT 07-MAY-2003 (first entry)

DE Helicobacter pylori selected interacting domain (STD) protein #711.

XX Protein-protein interaction; ulcer; selected interacting domain; STD.

OS Helicobacter pylori.

XX MO200266501-A2.

XX 29-AUG-2002.

XX 28-DEC-2001; 2001WO-BP015428.

XX 02-JAN-2001; 2001US-0259302P.

XX (HYBR-) HYBRIGENICS.

XX (INSP) INST PASTEUR.

XX Legrain P, Rain J, Colland F, De Reuse H, Labigne A;

XX WPI; 2002-674910/72.

XX N-PSDB; ABX6112.

XX Claim 6; Page 257; 642p; English.

CC The invention describes a complex of protein-protein interactions in
 CC Helicobacter pylori selected from 421 complexes given in the
 CC specification. The complex of protein-protein interactions are useful for
 CC screening for agents which modulate the interaction of proteins.
 CC Modulating compounds which binds to a targeted bacterial protein may be
 CC used for treating or preventing ulcers in a human or animal. This is the
 CC amino acid sequence of a selected interacting domain (STD), identified
 CC via protein-protein interactions. Note: Where the patent number printed
 CC at the top of the pages in the specification has obscured areas of
 CC protein sequence, the indexer has replaced the residue with an X to
 CC represent an illegible residue

SO Sequence 420 AA:

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	1,89e-82	420	193	2	20	3
Percent Similarity:	896.50					
Best Local Similarity:	77.4%					
Query Match:	76.6%					
DB:	76.6%					

US-09-360-685C-26 (1-685) x ABU51367 (1-420)

QY 37 CAAGCAAAAGCGACTTGAATTCGTTAAAGTGAATCATCAATCAAAAGGTACG 96
 DB 1 GlnAlaLysSerAspLysGlnAsnSerValLysAspValIleIleAsnGlnLysValThr 20
 QY 97 GATTAAGTTGATTAATCTCAATCAAGCGGTATCAAGTGAATCAAGCAAGCGGTGATTCAGT 156
 DB 21 AspLysValAspAsnLysAsnGlnAlaValSerValAlaLysAlaMetGlyAspPheSer 40
 QY 157 AGGTTAGAGCAAGCGGTAGCCGATCTCAAAAATTTCTCAAGAGCAATGTCACCAACAA 216

```

Db 41 ArgValGluGlnValLeuAlaSerPheLeuValSerPheSerLeuGlnLeuAlaGlnGln 60
Qy 217 GCTCAAAAATGAAATGCTCAATGCTAGAAAAAAATCTGAAATATCAATCCGTTAAG 276
Db 61 AlaGlnValSerPheLeuValSerPheLeuValSerPheLeuValSerPheLeuValSer 80
Qy 277 AATGCTGAAATGCAACCTTTCGTTAATGCTTATCTCAAGCAAGCAACCACTT 336
Db 81 AsnSerValAsnValSerPheLeuValSerPheLeuValSerPheLeuValSerPheLeu 100
Qy 337 TCTAAAACTTTTCGACATCAAGAAAGATGAAATGCAAACTTGAATTCATTAAC 396
Db 101 AlaValSerPheSerPheLeuValSerPheLeuValSerPheLeuValSerPheLeu 119
Qy 397 AATTAACATATGACTCAAAAAC-----GAAACCAATTAATGCTAATTAATTAAG 450
Db 120 AsnAsnAsnAsnGlyLeuValSerPheLeuValSerPheLeuValSerPheLeuValSer 139
Qy 451 AAAGCAAGGCAAGCAAGCTTACCTTGAAGAACCTTATACGCTCAAGTTGCTAAAGCA 510
Db 140 LysThrGlyGlnValAlaSerProGluGlnProLysThrGlnValAlaValSerVal 159
Qy 511 AATGCAAAATTTGACCGACTCAATCAATATGCAAGTGTGGTGTGTTAGCGCAAGCA 570
Db 160 AsnAlaValSerPheLeuValSerPheLeuValSerPheLeuValSerPheLeuValSer 179
Qy 571 GCGGGCTTCCTTGAAGAAAGCATATTAAGTTGATGATGATGATGATGATGATGATGAT 627
Db 180 AlaGlyPheProLeuValSerPheLeuValSerPheLeuValSerPheLeuValSer 199
Qy 627 ----- 627
Db 200 AlaSerProGluProLysThrAlaThrAlaSerPheLeuValSerPheProLeuValSer 219
Qy 628 ----- 628
Db 220 ArgHisAspValSerPheLeuValSerPheLeuValSerPheLeuValSerPheLeuValSer 239
Qy 649 CAGAAATTTGACATCTCAATCAAGCGCTTCAAGCA 684
Db 240 GlnValSerPheLeuValSerPheLeuValSerPheLeuValSerPheLeuValSer 251

```

RESULT 9
AEB91411
ID AEB91411 standard; protein; 1167 AA.
AC AEB91411;
XX
XX 20-OCT-2005 (first entry)
DT
XX
XX Microbial pathogen adhesin protein sequence. SEQ ID NO:121.
DE
XX
XX
KW algorithm; adhesin; pharmaceutical; vaccine; drug screening;
KW Bordetella pertussis infection; antibacterial; pneumonia;
KW antiinflammatory; respiratory-gen.; gastric ulcer; antifungal;
KW gastrointestinal-gen.; urinary tract infection; antimicrobial; uropathic.
OS Helicobacter pylori.
XX
XX PN WO2005076010-A2.
XX
XX PD 18-AUG-2005.
XX
XX PF 07-FEB-2005; 2005WO-IN00037.
XX
XX PR 06-FEB-2004; 2004IN-DE000173.
XX
XX PR 20-JUL-2004; 2004US-0589227P.
XX
XX PA (COUL.) COUNCIL SCI & IND RES SOUTH AFRICA.
XX Sachdeva G, Kumar K, Jain P, Brahmachari SK, Ramchandran S;

```

DR WPI, 2005-597835/61.
XX  
XX Computational method for identifying adhesin and adhesin like molecules,
XX comprises computing sequence-based attributes of protein sequences using
XX neutral network software and training an artificial neural network.
XX  
PS Claim 16; SEQ ID NO 121; 402pp; English.
XX  
XX The present invention relates to a computational method (M1) for
XX identifying adhesin and adhesin-like proteins, by computing the sequence-
XX based attributes of protein sequences using five attribute modules of a
XX neural network software, training an artificial neural network (ANN) for
XX each of the computed five attributes, and identifying the adhesin and
XX adhesin-like proteins having probability of being an adhesin (Pad) as
XX equal or greater than 0.51. Also claimed is a set of 274 annotated genes
XX encoding adhesin and adhesin-like proteins, having 274 fully defined
XX 1151 base pair (SEQ ID NO: 385-658) sequences; a set of 105 hypothetical
XX genes encoding adhesin and adhesin-like proteins, having 105 fully
XX defined 306-15876 base pairs (SEQ ID NO: 659-763), a set of 279 annotated
XX adhesin and adhesin-like proteins, having 279 fully defined 53-3716 base
XX pair (SEQ ID NO: 1-279); a set of 105 hypothetical adhesin and adhesin-
XX like proteins, having 105 fully defined 105-5291 base pair (SEQ ID NO:
XX 280-384) sequences; and a fully connected multilayer feed forward ANN (1)
XX based on (M1). (M1) is useful for identifying adhesin and adhesin-like
XX proteins, of therapeutic potential, and identifying adhesin and adhesin-like
XX proteins for further testing in development of new vaccine formulations
XX to eliminate diseases caused by various pathogenic organisms. (M1) is
XX useful for identifying putative adhesins that are important in drug
XX discovery and preventing therapeutic adhesins that are important in drug
XX gastric ulcer and urinary tract infections. (M1) identifies adhesins from
XX phylogenetic spectrum. (M1) is capable of predicting adhesive nature of
XX unique proteins. The present sequence is a microbial pathogen adhesin
XX protein sequence.
XX  
SQ Sequence 1167 AA;

```

Alignment Scores:

Pred. No.:	2,446-82	Length:	1167
Score:	896.50	Matches:	189
Percent Similarity:	74.88	Conservative:	7
Best Local Similarity:	72.18	Mismatches:	15
Query Match:	76.68	Indels:	51
DB:	9	Gaps:	2

US-09-360-685C-26 (1-685) x AEB91411 (1-1167)

```

Qy 1 AAAAAATGCAAAATTAAGATTTACAGATTAACGACAAAGCAAAAGCACTTGAAT 60
Db 754 LysAsnGlyValAsnValSerPheSerValThrGlnAlaValSerAspLeuGlnLeu 773
Qy 61 TCGCTTAAGATGATCATCAATCAAAAGATTAACGATTAAGTTGATTAATCTCAATCA 120
Db 774 SerLeuAspValAlaLeuValSerPheLeuValSerPheLeuValSerPheLeuValSer 793
Qy 121 GCGGTATCGTGGCTTAAGCAACGCGTATTTCAAGTGAAGGTAAGCAAGCTTACCGAT 180
Db 794 AlaValSerValAlaValSerPheSerValThrGlnAlaValSerPheLeuValSerPhe 813
Qy 181 CTCAAAATTTCTCAAAAGAGCAATTTGCCCAACCAAGCTCAAAAATTAAGTCTCAAT 240
Db 814 LeuValSerPheSerValSerPheLeuValSerPheLeuValSerPheLeuValSerPhe 833
Qy 241 GCTAGAAAAATCTGAAATATTAATCAATCGTTAAAGATGCTGAATGCAACCTTACTC 300
Db 834 ThrGlyValAsnSerValLeuValSerValValSerValValSerValValSerVal 853
Qy 301 GGTATGAGTTATCTCAAGCAAGCAAGCAACCTTCAAAAATTTTGGACATCAAG 360
Db 854 GlyAsnGlyLeuSerValAlaValSerValThrLeuSerValSerPheSerPheLeuValSer 873
Qy 361 AAAGAGTTGATGCAAAATCTGAAATTTTGAATTAATCAATTAATGATGATCAAAAG 420

```

Db 874 LysGlnLeuAsnAlaIleuGlyAsnPheAsnAsnAsnAsnGlnGluAsn 893
 Qy 421 GAACCATTTATGCTAAAGTTAAATAAAGAGGAGGAGAGCTACCTTGAAAGA 480
 Db 894 Ser-----ThnGlu 896
 Qy 481 CCCATTACGCTCAAGTTGCTAAAGTAATGCAAAATTCACGACTCAATCAATA 540
 Db 897 ProIleTyrThrGlnValAlaIleValIleValIleValIleValIleValIle 916
 Qy 541 GCAGAGTGTGGTGTGGTGTGGGCAAGCAGCGGC----- 576
 Db 917 AlaSerGlyLeuGlyAspValGlyGlnAlaAlaSerPheLeuLeuLysArgHisAspLys 936
 Qy 576 ----- 576
 Db 937 ValAspAspLeuSerLysValGlyLeuSerAlaAsnHisGlnProIleTyrAlaThrIle 956
 Qy 577 -----TTCCCTTTGAAAAAGCATGATTAAGTTGATTCAGTAAG 618
 Db 957 AspAspLeuGlyGlyProPheProLeuLysArgHisAspLysValAspAspLeuSerLys 976
 Qy 619 GTAGGCTTTCAAGCAATCAAGATTGCTCAGAAAAATTGACATCTCAATCAAGCGTA 678
 Db 977 ValGlyLeuSerArgGlnGlnLysLeuThrGlnLysIleAspAsnLeuGlnAlaVal 996
 Qy 679 TCAGAA 684
 Db 997 SerGln 998
 RESULT 10
 AAM20174
 ID AAM20174 standard; protein; 1178 AA.
 AC AAM20174;
 XX 08-JUL-1997 (first entry)
 DE H. pylori cytoplasmic protein, 16459375.aa.
 KM Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 OS duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 XX Helicobacter pylori.
 PN W09640893-A1.
 XX 19-DEC-1996.
 PD 06-JUN-1996; 96WO-US009122.
 PF 07-JUN-1995; 95US-00487032.
 PR 01-APR-1996; 96US-00630405.
 XX (ASTR) ASTRA AB.
 PA Smith D, Berglindh OT, Mellgaard BL;
 PI WPI; 1997-052306/05.
 DR N-PSDB; AAT67410.
 XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
 PT useful for vaccines to treat or prevent H. pylori infection, and to
 PT detect Helicobacter.
 PS Claim 61; Page 390-392; 1481pp; English.
 The present sequence is a Helicobacter pylori cytoplasmic protein. The
 protein may be used in a vaccine to prevent or treat H. pylori infection
 or to identify H. pylori polypeptide binding compounds, useful as
 potential H. pylori life cycle activators or inhibitors. The genomic
 sequence of H. pylori (ATCC 55679) was determined from overlapping

CC contigs generated by mechanically shearing the bacterial DNA. The
 CC sequences were analyzed for ORF of at least 180 nucleotides, and the
 CC predicted coding regions defined by computer evaluation. To identify
 CC likely H. pylori antigens for vaccine development, the amino acid
 CC sequences predicted from various ORF were analyzed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts
 XX
 SQ Sequence 1178 AA;
 Alignment Scores:
 Pred. No.: 2,45e-82 Length: 1178
 Score: 896.50 Matches: 189
 Percent Similarity: 74.8% Conservative: 7
 Best Local Similarity: 72.1% Mismatches: 15
 Query Match: 76.6% Indels: 51
 DB: 2 Gaps: 2
 US-09-360-685C-26 (1-685) x AAM20174 (1-1178)
 Qy 1 AAAAAATGCAAAATTAAGATTTCAGCAAGTAAAGCAAAAGCACTTGAATA 60
 Db 765 LysGlnGlyLysAsnLysAspPheSerLysValIleGlnAlaLysSerAspLeuGlnAsn 784
 Qy 61 TCCGTTAAAGATGTGATCATCATCAATCAAAAGTAAAGTAAAGTAAATTCATCAATCA 120
 Db 785 SerIleLysAspValIleIleAsnGlnLysIleThrAspLysValAspAsnLeuGln 804
 Qy 121 GCGGTATCAGTGGCTTAAAGCAACGGGTGATTCAGTAGGCTTAAAGCAAGCTTACCGAT 180
 Db 805 AlaValSerValAlaIleValAlaThrGlyAspPheSerGlyValGlnGlnAlaLeuAlaAsp 824
 Qy 181 CTCAAAAATTTCTCAAGAGCAATTGGCCCAACAGCAAGCTCAAAAAATGAAGTGCAT 240
 Db 825 LeuLysAsnPheSerLysGlnGlnLeuAlaGlnGlnAlaGlnLysAsnGlnAspPheAsn 844
 Qy 241 GCTAGAAAAATCTGAATATATATCATCCGTTAAGATGGTGTAATGAAACCTTACTC 300
 Db 845 ThrGlyLysAsnSerAlaLeuTyrGlnSerValLysAsnGlyValAsnGlyThrLeuVal 864
 Qy 301 GGTATGGGTATATCTCAAGCAAGCAACCAACTCTTTTAAAACTTTTCGACATCAAG 360
 Db 865 GlyAsnGlyLeuSerLysAlaGlnAlaThrThrLeuSerLysAsnPheSerAspIleLys 884
 Qy 361 AAGAGTTGAANTGCAAACTTGAAATTTCAATTAATCAATTAATGACTCAAAAC 420
 Db 885 LysGlnLeuAsnAlaLysLeuGlyAsnPheAsnAsnAsnAsnAsnGlyLeuGlnAsn 904
 Qy 421 GAACCATTTATGCTAAAGTTAAATAAAGAGGCAAGGCAAGCAGTACCTTGAGAA 480
 Db 905 Ser-----ThnGlu 907
 Qy 481 CCCATTACGCTCAAGTTGCTAAAGTAATGCAAAATTCACGACTCAATCAATA 540
 Db 908 ProIleTyrThrGlnValAlaIleValIleValIleValIleValIleValIle 927
 Qy 541 GCAGAGTGTGGTGTGGTGTGGGCAAGCAGCGGC----- 576
 Db 928 AlaSerGlyLeuGlyAspValGlyGlnAlaAlaSerPheLeuLeuLysArgHisAspLys 947
 Qy 576 ----- 576
 Db 948 ValAspAspLeuSerLysValGlyLeuSerAlaAsnHisGlnProIleTyrAlaThrIle 967
 Qy 577 -----TTCCCTTTGAAAAAGCATGATTAAGTTGATTCAGTAAG 618
 Db 968 AspAspLeuGlyGlyProPheProLeuLysArgHisAspLysValAspAspLeuSerLys 987
 Qy 619 GTAGGCTTTCAAGCAATCAAGATTGCTCAGAAAAATTGACATCTCAATCAAGCGTA 678
 Db 988 ValGlyLeuSerArgGlnGlnLysLeuThrGlnLysIleAspAsnLeuGlnAlaVal 1007

QY 679 TCAGAA 684
DB 1008 SerGlu 1009

RESULT 11

AAW20884
ID AAW20884 standard; protein; 1183 AA.

AAW20884;
18-JUL-1997 (first entry)

DE H. pylori cytoplasmic protein, 13ee12016orf74.

XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacterium; life cycle; activator;
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
diagnosis.

XX Helicobacter pylori.

XX WO9640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US009122.

XX 07-JUN-1995; 95US-00487032.

XX 01-APR-1996; 96US-00630405.

XX (ASTR) ASTRA AB.

XX Smith D, Berglindh OT, Mellgaerd BL;

XX WPI; 1997-052306/05.

XX N-PSDB; AAT68137.

XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
PT useful for vaccines to treat or prevent H. pylori infection, and to
PT detect Helicobacter.

XX Claim 61; Page 1285-87; 1481p; English.

XX This sequence represents a H. pylori cytoplasmic protein. The protein may
CC be used in a vaccine to prevent or treat H. pylori infection or to
CC identify H. pylori polypeptide binding compounds, useful as potential H.
CC pylori life cycle activators or inhibitors. The genomic sequence of H.
CC pylori (ATCC 55679) was determined from overlapping contigs generated by
CC mechanically shearing the bacterial DNA. The sequences were analysed for
CC ORF of at least 180 nucleotides, and the predicted coding regions defined
CC by computer evaluation. To identify likely H. pylori antigens for vaccine
CC development, the amino acid sequences predicted from various ORF were
CC analysed for significant homology to other known or exported membrane
CC proteins. Having identified and determined the sequences of interest
CC particular regions can be isolated from H. pylori by PCR amplification
CC for recombinant polypeptide production, e.g. in E. coli hosts

XX Sequence 1183 AA;

Alignment Scores:

Score: 2 45e-82 Length: 1183
Percent Similarity: 89e-50 Matches: 189
Best Local Similarity: 74.8% Conservative: 7
Query Match: 72.1% Mismatches: 15
Query Match: 76.6% Indels: 51
DB: 2 Gaps: 2

US-09-360-685C-26 (1-685) x AAW20884 (1-1183)

QY 1 AAAAATGGCAAAATTAAGATTTCAGCAAGCTTAAGCAAGCAAAAGCACTTGAAT 60
DB 770 LysAsnGlyLeuAsnLysAspPheSerLysValThrGlnAlaLysSerAspLeuGluAsn 789

QY 61 TCCGTTAAAGATGTGATCATCAATCAAAAGGTAAACGATTAAGTTGTAATCATCAATCA 120
DB 790 SerLeuLysAspValIleIleAsnGlnLysLeuThrAspLysValAspAsnLeuAsnGln 809
QY 121 GCGGTATGAGTGTAAAGCAACGGGTATTTCAGTGAAGGTGAAGCAAGCGTTAGCCAT 180
DB 810 AlaValSerValAlaLysAlaThrGlyAspPheSerGlyValGluGlnAlaLeuAlaAsp 829
QY 181 CTCAAAAATTTCTCAAAAGCAATTTGGCCCAACGACTCAAAATTAAGTCAATCAAT 240
DB 830 LeuLysAsnPheSerLysGluGlnLeuAlaGlnAlaGlnLysValAsnGlnLysPheAsn 849
QY 241 GCTAAGAAAAATCTGAAATATATCAATCCGTTAAGATGGTGAATGGAAACCCACTGC 300
DB 850 ThrGlyLysAsnSerAlaLeuLysGlnSerValLysAsnGlyValAsnGlyThrLeuVal 869
QY 301 GGTAAATGGTTATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 360
DB 870 GlyAsnGlyLeuSerLysAlaGlnAlaThrThrLeuSerLysAsnPheSerAspLys 889
QY 361 AAAGAGTTGAATGCAAACTTGAATTTCAATTAACATTAATTAATGACTCAAAAC 420
DB 890 LysGluLeuAsnAlaLysLeuGlyAsnPheAsnAsnAsnAsnAsnGlyLeuGluAsn 909
QY 421 GAACCCATTTATGCTAAAGTTAATTAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAG 480
DB 910 Ser-----ThrGlu 912
QY 481 CCCATTACGCTCAAGTGTCTTAAAGGTAAATGCAAAATTAAGCAAGCAAGCAAGCAAT 540
DB 913 ProLeuLysThrGlnValAlaLysLeuValLysValLysValLysValLysValLys 932
QY 541 GCAAGTGTGTTGGTGTGTGAAGGCAAGCAAGCGGCGC----- 576
DB 933 AlaSerGlyLeuGlyAspValGlyGlnAlaLysPheLeuLeuLysArgHisAspLys 952
QY 576 ----- 576
DB 953 ValAspAspLeuSerLysValGlyLeuSerAlaAsnHisGluProlLeuLysAlaThrIle 972
QY 577 -----TTCCCTTGAAGAGGCAATGAATGAAGTTGATGCTCAAGTAA 618
DB 973 AspAspLeuGlyGlyProPheProLeuLysArgHisAspLysValAspAspLeuSerLys 992
QY 619 GTAGGCTTGAAGAAATCAAGAAATGGCTCAAGAAATTTGCAATTCATCAACGGTGA 678
DB 993 ValGlyLeuSerAspGluGlnLysLeuThrGlnLysIleAspAsnLeuAsnGlnAlaVal 1012
QY 679 TCAGAA 684
DB 1013 SerGlu 1014

RESULT 12

ABU51383
ID ABU51383 standard; protein; 389 AA.

ABU51383;
07-MAY-2003 (first entry)

DE Helicobacter pylori selected interacting domain (SID) protein #727.

KW Protein-protein interaction; ulcer; selected interacting domain; SID.

OS Helicobacter pylori.

PN WO200266501-A2.

PD 29-AUG-2002.

PF 28-DEC-2001; 2001WO-EP015428.

US-09-360-685C-26 (1-685) x ABUS1625 (1-356)

```

QY 160 GTAGAGCAGCGCTTGAAGCAATTTCTCAAGAGCAATGGCCCAACAGCT 219
Db 1 ValGlnGlnValLeuAlaAspLeuLysAsnPhenSerLysGlnGlnLeuAlaGlnGlnAla 20
QY 220 CAAAAAATGAAAGCTCAATGCTAGAAAAAATGAAATATATCAATCCGTTAAGAAAT 279
Db 21 GlnLysAsnGlnAspPheAsnThrGlyLysAsnSerGlnLeuLysGlnSerValLysAsn 40
QY 280 GGTTGTAATGAAACCTTGCTGTAATGGTTATCTCAAGCAAGCAACCACTTTCT 339
Db 41 SerValAsnLysThrLeuValGlnLysGlnLysSerGlyLysGlnAlaThrAlaLeuAla 60
QY 340 AAAAAGCTTTGGCAGATCAAGAAAAGTTGAATGCAAACTTGAAATTTCAATTAACAT 399
Db 61 LysAsnPhenSerAspLysLysGlnLeuAsnGlnLysPheLysAsnPhe---AsnAsn 79
QY 400 AACAAATATGAGCTCAAAAAC-----GAACCAATTTATGCTAAAGTTAATAAAGAAA 453
Db 80 AsnAsnAsnGlnLysLeuLysAsnSerThrGlnProLysThrAlaLysValAsnLysLysLys 99
QY 454 GCAAGGCAAGCAGCTAGCTTGAAGAACCCATTATGAGCTCAAGTTGCTAAAGGTAAT 513
Db 100 ThrGlnGlnValAlaSerProGlnGlnProLysThrGlnValAlaLysLysValAsn 119
QY 514 GCAAAAATTTGACGAGCTCAATCAATAGCAAGTGGTTGGGTGTAAGGCGAAGCAG 573
Db 120 AlaLysIleAspArgLeuAsnGlnIleLysSerGlnLysGlnLysGlnAlaAla 139
QY 574 GGCTTCCTTTGAAAAGCATGATTAAGTTGATGATCTCAAGTAAAGAGAGGCTT----- 627
Db 140 GlyPheProLeuLysArgHisAspLysValAspAspLeuSerLysValGlnLeuSerAla 159
QY 627 ----- 627
Db 160 SerProGlnProLysThrAlaThrIleAspAspLeuGlnGlyProPheProLeuLysArg 179
QY 628 -----TCAGGAATCAAGAAATGGCTCAG 651
Db 180 HisAspLysValAspAspLeuSerLysValGlnLysSerArgAsnGlnLeuAlaGln 199
QY 652 AAAATGCAATCTCAATCAAGCGTATCGAA 684
Db 200 LysIleAspAsnLeuAsnGlnAlaValSerGln 210

RESULT 14
ABUS0882
ID ABUS0882 standard; protein; 379 AA.
AC ABUS0882;
XX
XX 07-MAY-2003 (first entry)
XX
XX Helicobacter pylori selected interacting domain (SID) protein #225.
XX DE Helicobacter pylori selected interacting domain (SID) protein #225.
XX KM Protein-protein interaction; ulcer; selected interacting domain; SID.
XX OS Helicobacter pylori.
XX
XX WO200266501-A2.
XX
XX 29-AUG-2002.
XX
XX 28-DEC-2001; 2001WO-EP015428.
XX PF
XX 02-JAN-2001; 2001US-0259302P.
XX PR
XX (HYBR-) HYBRIGENICS.
XX PA (INSP) INST PASTEUR.
XX
XX Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
XX

```

DR MPI: 2002-674910/72.

DR N-PSDB; ABX65626.

XX
XX
XX New complexes of protein-protein interactions in Helicobacter pylori,
XX useful for identifying modulating compounds for treating or preventing
XX ulcers in mammals.

XX Claim 6; Page 140; 642pp; English.

XX The invention describes a complex of protein-protein interactions in
XX Helicobacter pylori selected from 421 complexes given in the
XX specification. The complex of protein-protein interactions are useful for
XX screening for agents which modulate the interaction of proteins.
XX Modulating compounds which binds to a targeted bacterial protein may be
XX used for treating or preventing ulcers in a human or animal. This is the
XX amino acid sequence of a selected interacting domain (SID), identified
XX via protein-protein interactions. Note: Where the patent number printed
XX at the top of the pages in the specification has obscured areas of
XX protein sequence, the index has replaced the residue with an X to
XX represent an illegible residue

XX Sequence 379 AA;

Alignment Scores:

Pred. No.:	7,346-63	Length:	379
Score:	704.50	Matches:	153
Percent Similarity:	73.5%	Conservative:	2
Best Local Similarity:	72.5%	Mismatches:	19
Query Match:	60.2%	Indels:	37
DB:	5	Gaps:	3

US-09-360-685C-26 (1-685) x ABUS0882 (1-379)

```

QY 160 GTAGAGCAGCGCTTGAAGCAATTTCTCAAGAGCAATGGCCCAACAGCT 219
Db 1 ValGlnGlnValLeuAlaAspLeuLysAsnPhenSerLysGlnGlnLeuAlaGlnGlnAla 20
QY 220 CAAAAAATGAAAGCTCAATGCTAGAAAAAATGAAATATATCAATCCGTTAAGAAAT 279
Db 21 GlnLysAsnGlnAspPheAsnThrGlyLysAsnSerGlnLeuLysGlnSerValLysAsn 40
QY 280 GGTTGTAATGAAACCTTGCTGTAATGGTTATCTCAAGCAAGCAACCACTTTCT 339
Db 41 SerValAsnLysThrLeuValGlnLysGlnLysSerGlyLysGlnAlaThrAlaLeuAla 60
QY 340 AAAAAGCTTTGGCAGATCAAGAAAAGTTGAATGCAAACTTGAAATTTCAATTAACAT 399
Db 61 LysAsnPhenSerAspLysLysGlnLeuAsnGlnLysPheLysAsnPhe---AsnAsn 79
QY 400 AACAAATATGAGCTCAAAAAC-----GAACCAATTTATGCTAAAGTTAATAAAGAAA 453
Db 80 AsnAsnAsnGlnLysLeuLysAsnSerThrGlnProLysThrAlaLysValAsnLysLysLys 99
QY 454 GCAAGGCAAGCAGCTAGCTTGAAGAACCCATTATGAGCTCAAGTTGCTAAAGGTAAT 513
Db 100 ThrGlnGlnValAlaSerProGlnGlnProLysThrGlnValAlaLysLysValAsn 119
QY 514 GCAAAAATTTGACGAGCTCAATCAATAGCAAGTGGTTGGGTGTAAGGCGAAGCAG 573
Db 120 AlaLysIleAspArgLeuAsnGlnIleLysSerGlnLysGlnLysGlnAlaAla 139
QY 574 GGCTTCCTTTGAAAAGCATGATTAAGTTGATGATCTCAAGTAAAGAGAGGCTT----- 627
Db 140 GlyPheProLeuLysArgHisAspLysValAspAspLeuSerLysValGlnLeuSerAla 159
QY 627 ----- 627
Db 160 SerProGlnProLysThrAlaThrIleAspAspLeuGlnGlyProPheProLeuLysArg 179
QY 628 -----TCAGGAATCAAGAAATGGCTCAG 651
Db 180 HisAspLysValAspAspLeuSerLysValGlnLysSerArgAsnGlnLeuAlaGln 199

```

QY 652 AAAATTGACATCTCAATCAGCGGTATTCAGAA 684
 DB 200 LysileaspAsnleuEnglnAlaValSerGlu 210

RESULT 15

ID ABUS1953 standard; protein; 284 AA.

AC ABUS1953;

DT 08-MAY-2003 (first entry)

DE Helicobacter pylori selected interacting domain (SID) protein #1297.

KM Protein-protein interaction; ulcer; selected interacting domain; SID.

XX Helicobacter pylori.

PN WO200266501-A2.

PD 29-AUG-2002.

PF 28-DEC-2001; 2001WO-EP015428.

PR 02-JAN-2001; 2001US-0259302P.

PA (HYBR-) HYBRIGENICS.

PA (INSP) INST PASTEUR.

PI Legrain P, Rain J, Colland F, De Reuse H, Labigne A;

DR WPI; 2002-674910/72.

DR N-PSDB; ABX66698.

PT New complexes of protein-protein interactions in Helicobacter pylori,
 useful for identifying modulating compounds for treating or preventing
 ulcers in mammals.

PS Claim 6; Page 399; 642pp; English.

CC The invention describes a complex of protein-protein interactions in
 CC Helicobacter pylori selected from 421 complexes given in the
 CC specification. The complex of protein-protein interactions are useful for
 CC screening for agents which modulate the interaction of proteins.
 CC Modulating compounds which binds to a targeted bacterial protein may be
 CC used for treating or preventing ulcers in a human or animal. This is the
 CC amino acid sequence of a selected interacting domain (SID), identified
 CC via protein-protein interactions. Note: Where the patent number printed
 CC at the top of the pages in the specification has obscured areas of
 CC protein sequence, the indexer has replaced the residue with an X to
 CC represent an illegible residue

XX Sequence 284 AA;

Alignment Scores:

Pred. No.:	2 21e-62	Length:	284
Score:	699.50	Matches:	152
Percent Similarity:	73.0%	Conservative:	2
Best Local Similarity:	72.0%	Mismatches:	20
Query Match:	59.7%	Indels:	37
DB:	5	Gaps:	3

US-09-360-685c-26 (1-685) x ABUS1953 (1-284)

QY 160 GTAGAGCAGCGGTATTCAGAAATTTCTCAAGAGCAATGGCCCAACAGCT 219
 DB 1 ValGluGlnValLeuAlaAspLeuLysAsnPheserIySgluGlnleuAlaGlnAla 20
 QY 220 CAAAAAATGAAAGTCTCAATGCTAGAAAAAATCTGAATATATCAATCGTTAAGAT 279
 DB 21 GlnIyAsnGlnuAspPheserThrGlyIyAsnSerGlnLeuIyGlnSerValIyAsn 40
 QY 280 GGTGTGATGAAACCTTAGTGTGATATGCTCAAGAGCAAGCAACACTCTTCT 339

DB 41 SerValAsnIySThrLeuValGlyAsnGlyLeuSerGlyIleGlnAlaThrAlaLeuAla 60
 QY 340 AAAAACTTTTCGACATCAGAAAGATTGAAATGCAAACTTGAATTTCAATTAACAAT 399
 DB 61 LysAsnPheserAspIleIySgluGlnleuEnglnIySPhenIyAsnPheserAsn 79
 QY 400 AACATTAATGACTCAAAAAC-----GAACCATTTATGCTAAAGTTAATTAAGAA 453
 DB 80 AsnAsnAsnGlyLeuLysAsnSerThrGlnProIleTyrAlaIySValAsnIyS 99
 QY 454 GCAGGCAAGCAGCTTAGCCTTGAAGACCAATTTAGCTCAAGTGTCTAAAGTAAAT 513
 DB 100 ThrGlyGlnValAlaSerProGlnuProIleTyrThrGlnValAlaIySValAsn 119
 QY 514 GCAAAATTAAGCCGACTCAATCAATCAATAGCAAGTGTGGTGTGTGAGGCAAGCAG 573
 DB 120 AlalysIleAspAspArgLeuAsnGlnIleAlaSerGlyLeuGlyGly**GlyGlnAla 139
 QY 574 GGCTTCCTTTGAAAAAGCATGATTAAGTTGATGATCTCAATTAAGTGGCTT----- 627
 DB 140 GlyPheProLeuIySArgHisAspIySValAspIySLeuSerIySValGlyLeuSerAla 159
 QY 627 ----- 627
 DB 160 SerProGlnuProIleTyrAlaThrIleAspAspLeuGlyGlyProPheProLeuIySArg 179
 QY 628 -----TCAGAGAAATCAAGAAATGGCTCAG 651
 DB 180 HisAspIySValAspAspLeuSerIyS**GlyIySserArgAsnGlnleuAlaGln 199
 QY 652 AAAATTGACATCTCAATCAGCGGTATTCAGAA 684
 DB 200 LysileaspAsnleuEnglnAlaValSerGlu 210

Search completed: March 6; 2006, 19:53:11
 Job time : 166 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2006, 21:29:20 ; Search time 3706 Seconds

(without alignments)
8647.909 Million cell updates/sec

Title: US-09-360-685C-26

Perfect score: 685

Sequence: 1 aaaaatgcaacaataagga.....caatcaagcgatcagaag 685

Scoring table: IDENTITY_NIC

Gapop 10.0, Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: gb_est1:
2: gb_est2:
3: gb_est3:
4: gb_hic:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_est7:
9: gb_gse1:
10: gb_gse2:
11: gb_gse3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60.4	8.8	1362	10	AG429915 Mus muscu
2	59.8	8.7	747	10	AL100640 Drosophila
3	59	8.6	1132	9	CC248307 CH261-881
4	58.6	8.6	1007	11	CNS06X9S
5	57.6	8.4	1390	10	AG347224
6	57.2	8.4	1247	8	DN698953
7	57	8.3	954	5	BX362982
8	56.2	8.2	1239	10	AG305325
9	55.6	8.2	1366	8	DN718180
10	55.6	8.1	1400	10	AG430154
11	54.8	8.0	1364	8	DN693282
12	54.6	8.0	1071	1	AJ926560
13	54.6	8.0	1128	1	CG753708
14	54.4	7.9	1223	10	CL103754
15	54.4	7.9	1266	10	AJ859745
16	54.4	7.9	1475	10	CL495869
17	54.2	7.9	1104	1	AJ925669
18	54	7.9	1435	10	AG347018
19	53.8	7.9	802	10	AJ885252
20	53.8	7.9	971	6	CD388253
21	53.8	7.9	1101	10	CNS00DG1
22	53.6	7.8	1028	8	DN564006

23	53.6	7.8	1101	10	CNS0182P	AL108811 Drosophila
24	53.6	7.8	1399	10	AG311276	Mus muscu
25	53.4	7.8	907	10	CNS021U4	AL176953 Tetradon
26	53.4	7.8	976	10	AG136173	AG136173 Pan trogl
27	53	7.7	1137	10	CL080345	CL080345 CH216-158
28	52.8	7.7	1335	10	CL647352	CL647352 CH213-136
29	52.8	7.7	1928	8	DR154244	DR154244 49132583
30	52.6	7.7	834	9	B12387	B12387 F21E20-Spc.
31	52.6	7.7	1101	10	CNS017KX	AL108171 Drosophila
32	52.6	7.7	1401	10	AG346504	AG346504 Mus muscu
33	52.2	7.6	783	10	CNS011PP	AL100567 Drosophila
34	52.2	7.6	1173	10	AJ926089	DN705530 AJ926089
35	52	7.6	1330	8	DN705530	DN705530 CL360-110
36	52	7.6	1640	10	AG393452	AG393452 Mus muscu
37	51.8	7.6	1195	8	DR146887	DR146887 49090646
38	51.8	7.6	1201	9	BZ570849	BZ570849 msh2 1563
39	51.8	7.6	1221	9	CC301561	CC301561 CH261-13K
40	51.8	7.6	1683	10	CL075415	CL075415 CH216-136
41	51.6	7.5	804	11	CNS03C75	AL237434 Tetradon
42	51.4	7.5	661	9	AQ378322	AQ378322 RPI111-16
43	51.4	7.5	1227	10	CG746198	CG746198 P039-2-B0
44	51.4	7.5	1269	10	CG749445	CG749445 P043-3-G0
45	51.4	7.5	1291	10	AG387132	AG387132 Mus muscu

ALIGNMENTS

RESULT 1
AG429915/c
LOCUS
DEFINITION
Mus musculus molossinus DNA, clone:MSMg01-304L12.TU, genomic survey
sequence.
ACCESSION
AG429915
VERSION
AG429915.1
KEYWORDS
GSS.
SOURCE
Mus musculus molossinus (Japanese wild mouse)
ORGANISM
Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1
Abe, K., Noguchi, H., Tagawa, K., Yuzuriha, M., Toyoda, A., Kojima, T.,
Ezawa, K., Saitou, N., Hattori, M., Sakaki, Y., Moriwaki, K. and
Shirolahi, T.
Contribution of Asian mouse subspecies Mus musculus molossinus to
genomic constitution of strain C57BL/6J, as defined by BAC-end
sequence-SNP analysis
Genome Res. 14 (12), 2439-2447 (2004)
15574823
2 (bases 1 to 1362)
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
Direct Submission
Submitted (17-NOV-2003) Masmahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Saitoh-cho, Tsukuba, Ibaraki, Japan, 305-0856, Japan
(E-mail: hattori@gsr.riken.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the mouse BAC library MSMG01. For BAC
library availability, please contact Kuniya Abe (abe@rc.riken.jp).
Takubata Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rc.riken.jp
PRIMERS
Sequencing : TU
Library : PBACE3.6
Vector :
R.site 1 : EORI
R.site 2 : EORI
Location/Qualifiers
1. 1362

Seq primer: RM1 TACGACTCACTATAGGAGAGA
 Class: BAC ends
 High quality sequence start: 28
 High quality sequence stop: 165.
 Location/Qualifiers

FEATURES

Source

1. 1132
 /organism="Gallus gallus"
 /mol_type="genomic DNA"
 /strain="Red Jungle Fowl"
 /db_xref="taxon:9031"
 /clone="CH261-8817"
 /sex="female"
 /cell_line="UCD001, indred 256"
 /clone_1ib="CH261"
 /note="Vector: PTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
 CH261 Female Chicken library - for library and clone
 ordering information: http://www.choitl.org/bacpac"

ORIGIN

Query Match 8.6%; Score 59; DB 9; Length 1132;
 Best Local Similarity 37.3%; Pred. No. 0.0011;
 Matches 203; Conservative 0; Mismatches 341; Indels 0; Gaps 0;

```

QY 1 AAAAAATGCGCAAAATATAGGATTTCAAGCAGTACGCAAGCAAAAGCACTTGAAT 60
DB 1015 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 956
QY 61 TCCGTTAAAGATGTGATCATCAATCAAAGGTAACGATTAAGTGAATCTCAATCA 120
DB 955 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 896
QY 121 GCGGTATCAAGTGTAAAGCAAGGATTTCAAGTGTAAAGCAAGGTTAGCCGAT 180
DB 895 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 836
QY 181 CTCAAAATTTCTCAAGAGCAATTTGCGCCCAAGCTCAAAAATATGAACTCAAT 240
DB 835 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 776
QY 241 GCTAGAAAAAATCTGAATATATCAATCCGTTAAGATGTGTGAATGAACTTAC 300
DB 775 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 716
QY 301 GGTATGAGGTATCTCAAGCAAGCAAGCTTTCTTAAATCTTTCGACATCAAG 360
DB 715 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 656
QY 361 AAGAGTTGATGCAAACTTGAAATTTCAATTAACAATTAACAATTAACAATTA 420
DB 655 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 596
QY 421 GAACCAATTTATGCTAAAGTAAATGAAGAAAGCAAGGCAAGCAAGCTTGAAGA 480
DB 595 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 536
QY 481 CCCATTAGCGCTCAAGTGTAAAGGTAATGCAAAATTAACGACCTCAATCAATA 540
DB 535 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 476
QY 541 GCAG 544
DB 475 AAAA 472

```

RESULT 4
 CDS06X9S 1007 bp DNA linear GSS 06-UTL-2001
 LOCUS T3 end of clone AX0AA039F08 of library AX0A from strain CBS 7064
 DEFINITION of Pichia farinosa, genomic survey sequence.
 ACCESSION AL419462
 VERSION AL419462.1 GI:12202640
 KEYWORDS GSS.
 SOURCE Pichia farinosa

ORGANISM

Pichia farinosa

REFERENCE

1 (bases 1 to 1007)
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bolotin-Fukuhara,M., Bon,E., Brotier,P., Casargola,S.,
 deMontigny,J., Dujon,B., Durrens,P., Legingle,A., Liorente,B.,
 Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 FEBS Lett. 487 (1), 3-12 (2000)

TITLE

2 (bases 1 to 1007)
 de Montigny,J., Spehner,C., Souciet,J., Tekala,F., Dujon,B.,
 Wincker,P., Artiguenave,F. and Potier,S.
 Genomic exploration of the hemiascomycetous yeasts: 15. Pichia
 sorbitophila
 FEBS Lett. 487 (1), 87-90 (2000)

REFERENCE

11152876
 3 (bases 1 to 1007)
 Genoscope.
 Direct Submission
 Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage,
 2 rue Gaston Crémieux, CP 5706, 91057 Evry cedex, FRANCE. (E-mail :
 segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxi,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

JOURNAL

11152890
 3 (bases 1 to 1007)
 Genoscope.
 Direct Submission
 Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage,
 2 rue Gaston Crémieux, CP 5706, 91057 Evry cedex, FRANCE. (E-mail :
 segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxi,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

COMMENT

1. 1007
 /organism="Pichia farinosa"
 /mol_type="genomic DNA"
 /strain="CBS 7064"
 /db_xref="taxon:4920"
 /clone="AX0AA039F08"
 /clone_1ib="AX0AA"
 /note="end : 73"

FEATURES

source

ORIGIN

Query Match 8.6%; Score 58.6; DB 11; Length 1007;
 Best Local Similarity 37.1%; Pred. No. 0.0014;
 Matches 209; Conservative 62; Mismatches 286; Indels 6; Gaps 1;

```

QY 1 AAAAAATGCGCAAAATATAGGATTTCAAGCAGTACGCAAGCAAAAGCACTTGAAT 60
DB 105 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 164
QY 61 TCCGTTAAAGATGTGATCATCAATCAAAGGTAACGATTAAGTGAATCTCAATCA 120
DB 165 NAAAAAAAAATTAATTTTAAAAAGAAAGAAATGAAATTAATAAATAAATAAATAA 224
QY 121 GCGGTATCAAGTGTAAAGCAAGGATTTCAAGTGTAAAGCAAGGTTAGCCGAT 180
DB 225 GTRCATGTWCMWGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 284
QY 181 CTCAAAATTTCTCAAGAGCAATTTGCGCCCAAGCTCAAAAATATGAAAGTCAAT 240
DB 285 AAAAAAAAAAGAGAGGTTATTAATAAATAAATAAATAAATAAATAAATAAATAA 344
QY 241 GCTAGAAAAAATCTGAATATATCAATCCGTTA-----GAATGCTGTGAATGAACC 294
DB 345 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 404
QY 295 CTAGTCGATGATGATGATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 354

```

[illegible]

RESULT 5	AG347224/C	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
	AG347224	1390 bp	DNA					
	Mus musculus							
	molossinus							
	DNA, clone:MM901-14X18.T7,							
	sequence.							
	AG347224							
	AG347224.1							
	GI:47920534							
	Mus musculus							
	molossinus							
	Mus musculus							
	molossinus							
	(Japanese wild mouse)							

ORGANISM (Japanese wild mouse)
Mus musculus molossinus
Mus musculus molossinus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Murinae; Mus.
 1

TITLE	Ezama, R., Saitou, N., Hattori, M., Sakaki, Y., Moriwaiki, K. and Shiroishi, T.
JOURNAL	Contribution of Asian mouse subspecies <i>Mus musculus molossinus</i> to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
PUBLISHED	Genome Res. 14 (12), 2439-2447 (2004)
	15574823

JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
 PUBMED 15574823
 REFERENCE 2 (bases 1 to 1390)
 AUTHORS Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (17-NOV-2003) Masahito Hattori, The Institute of

COMMENT

and Chemical Research (RIKEN), The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045, Japan; E-mail: hatori@riken.jp, URL: <http://hgs.gsc.riken.go.jp>; Tel: 81-45-503-9191, Fax: 81-45-503-9170. Clones are derived from the mouse BAC library MSWG01. For BAC library availability, please contact Kunyia Abe (abe@riken.jp). The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Tsukuba Institute, Tsukuba, 305-0874 Japan. E-mail: abe@riken.jp; Phone: 81-298-36-9189; Fax: 81-298-36-9199

```

Sequencing : T7
LIBRARY
Vector      : pBAC3.6
R.Site 1    : EcoRI
R.Site 2    : EcoRI
FEATURES
    location/Qualifiers
        source
            1..1390

```

```

/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-142K18.77"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC library"

```

ORIGIN	
Query Match	8.4%
Best Local Similarity	44.4%
Ped. No. 0.0026;	
Matches 240; Conservative	0; Mismatches 304; Indels 0; Gaps 0

```

Qy 1 AAAAAATGCAAAAATAAGATTCTTCAGCAAGGTAGCAAGCAAAAAAGCACTTGAAT 60
   |||||
Db 1072 AAAAAATGCAAAAATAAGATTCTTCAGCAAGGTAGCAAGCAAAAAAGCACTTGAAT 1013
   |||||
Qy 61 TCCGTTAAAGATGATCATCATCATCAAAAGGTACGGATTAAAGTTGATTATCTCATCA 120
   |||||
Db 1012 GTAAAAAATAATATGAGATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 953
   |||||
Qy 121 GGGGTATCGGTGGCTTAAGCAAGGGGTATTTCAGTGGGTAGAGCAAGCCTTAGCGAT 180
   |||||
Db 952 ATAAAAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 893
   |||||
Qy 181 CTCAAAAATTTCTCAAGAGCAATGGCCCAACAAGCTCAAAAAATAAGAACTCAAT 240
   |||||
Db 892 AGAAAGAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 833
   |||||
Qy 241 GCTGAAAAAATCTGAATATATCAATCCGTTAAGATGGTGTAGTGAAGCAAGCCTAGTC 300
   |||||
Db 832 AATTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 773
   |||||
Qy 301 GGTATGGGTTATCTCAAGCAAGCAACAAGCTTTCTAAAACTTTGGCAATCAAG 360
   |||||
Db 772 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 713
   |||||
Qy 361 AAGAGATGCAATGCAAATCTGAAATTTCAATATACATATACATATGAGCTCAAAAC 420
   |||||
Db 712 AATTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 653
   |||||
Qy 421 GAACCATTTATGCTAAAGTTAATAAAGAAACAGAGGCAAGCTAGCCTTAGACA 480
   |||||
Db 652 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 593
   |||||
Qy 481 CCAATTTAGCTAGTGTCTAAAGATTAATGCAAAATATGACCAATCATCAATA 540
   |||||
Db 592 AAAAAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 533
   |||||
Qy 541 GCAA 544
   |||||
Db 532 AAAA 529

```

RESULT 6	
DN698953/C	
LOCUS	1247 bp
DEFINITION	mRNA
ACCESSION	EST 30-MAR-2005
VERSION	
KEYWORDS	ClJ20-F03 5' mRNA sequence.
	DN698953
	DN698953.1
	GI:62060731
	EST.

ORGANISM	REFERENCE AUTHORS	TITLE
Gasterosteus aculeatus (livee grained stickleback)	Klingesley, D.M., Felchak, C., Balabhadra, S., Grimwood, J., Dickson, M., Schmutz, J., and Myers, R.M.	Expressed sequence tags from Gasterosteus aculeatus
Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorphi; Acanthopterygii; Percomorpha; Gasterosteiformes; Gasterosteidae; Gasterosteus. 1 (bases 1 to 1247)	Unpublished (2003) Contact: Grimwood, Jane	

Stanford Human Genome Center
Stanford University School of Medicine
975 S California Ave, Palo Alto, USA
Tel: 650 320 5917
Fax: 650 320 5801
Email: jane@hngc.stanford.edu

D_b 216 AAAAAAAAAAGCCATAAAAAAAAAAAAAAAAAAAAAAAAA 157
O_y 512 ATGCGAAATTTACGACCTAATCAATAGCA 544
D_b 156 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 124

RESULT 8
AG305325/C

LOCUS			
DEFINITION	AG305325	1329 bp	DNA linear GSS 18-DEC-2004
ACCESSION	Mus musculus molossinus		
VERSION	sequence.		
KEYWORDS	AG305325		
SOURCE	AG305325.1	GI:47678279	GSS.
ORGANISM	Mus musculus molossinus		
	Mus musculus molossinus	(japanese wild mouse)	
	Mus musculus molossinus		

REFERENCE AUTHORS

TITLE	JOURNAL	REFERENCE
Contribution of Asian mouse subpopulations to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis	Genome Res.	14 (12), 2439-2447 (2004)
15574823		
2 (baaes 1 to 1239)		

TITLE
JOURNAL

1-7-22 Suehiro-chou, Tsutsumi-ku, Yokohama, Kanagawa, 230-0045, Japan
(E-mail: hatacori@ccg.riken.jp, URL: <http://kgp.ccg.riken.go.jp/>,
Tel: 81-45-503-9111, Fax: 81-45-503-9110)
Clones are derived from the mouse *3T3* (JCR-80, Wistar).

COMMENT

library availability, please contact Kuniya Abe (abe@rc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyda, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rc.riken.jp
prw@rc.riken.jp

FEATURES

```

/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSK901-085N20.17"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSK901 Mouse Male BAC Library"

```

ORIGIN

Query Match	8.2%	Score	56.2;	DB	10;	Length	1239;
Best local Similarity	43.6%						

Matches

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

 γ

1078 AAGGAGAGAAAAATATATAGAGAGCAAGAGAGAGAAAAAGTAGAAGATATATAGTC 1019

Qy	122	CGGTATCATGCGCTTAAGCAACGGGTGATTTCTAGTAGGATGAGCAAGCGTTAGCCGATC	181
Db	1018	ATTATTAAGAGATTAATTTAAAGAAAAATAAAGAAAGATGAGATTAATATGAAG	955
Qy	182	TCAAAAATTTCTCAAGAGCAATTGCGCCACAGACTCAAAAAATGAATGTCTAATG	241
Db	958	AAATATATTAATTAAGATTAAGATTAATTAAGATTAAGATTAAGATTAAGATTAAG	899
Qy	242	CTAGAAAAATCTGAATATATCATTCCTTAAGATGCTGTGATGAAACCTGTGCG	301
Db	898	GAGAAAGAAAAAAGAAATTAAGAGGGGATAGAAAGAGAGAAAAAAGTAT	839
Qy	302	GTATGCGTTATCTCAGCAGAACCAACTCTTTCTAAAACTTTTCGAGCATCAGA	361
Db	838	AAAGGGGAAAAAATTAATAATTAAGAAAGAAAGAAAGAAAGAAAGTAA	779
Qy	362	AGAATTGAATSCAAACTTGGAATTTCAATTAACAATTAATGATGACTAAAAAG	421
Db	778	AAAGATAGATTAATTAAGAGAAAAAAGAAAAAAGAAAGAAATTAAGAAAGAG	719
Qy	422	AACCATTTATGCTAAAGTTAATAAAGAAAGCGGCAAGCACTGACCTTGAAAGC	481
Db	718	AAAAAAGAAAAAGAAAGCAAGAAAGAAAGAAAGAAAGAAAGTAAAAAAGAAAAA	659
Qy	482	CCATTTACGCTCAAGTTGCTAAAAAGTTAAATGCAAAAATTAACGACTCATCAATAG	541
Db	658	GAAGAAAGAAATAGCATTAAGCAGGAGAGAAAAAAGAAAGAAAGAAAGTAA	599
Qy	542	CNA 544	
Db	598	GAA 596	

LOCUS

DEFINITION	1366 bp mRNA linear EST 31-MAR-2002
ACCESSION	CBN121-H10.y1d-B SHCC-CNB <i>Gasterosteus aculeatus</i> cDNA clone
VERSION	DN718180
KEYWORDS	DN718180.1 GI:62094416
SOURCE	EST.
ORGANISM	<i>Gasterosteus aculeatus</i> (three spined stickleback)

AUTHORS

TITLE Expressed sequence tags from *Gasterosteus aculeatus*
JOURNAL Unpublished (2003)
COMMENT Contact: Grimwood, Jane
Stanford Human Genome Center
Stanford University School of Medicine
975 S California Ave, Palo Alto, CA 94304, USA
Tel: 650 320 5917
Fax: 650 320 5801

email: jane@ngc.stanford.edu
plate: 121
High quality sequence stop: 814.

Source

```

/mol_type="mRNA"
/strain="Cormor Creek sticklebacks, WA USA"
/db_xref="taxon:69293"
/clone="CMB121-H10"
/sex="mixed male and female"
/tissue_type="Brain"
/dev_stage="adult"
/lab_host="DH10B (T1 phage resistant)"

```

/clone.lib="SHGC-CNB"
 /note="Vector: Express 1; Total and poly A+ RNA was isolated from the indicated stickleback tissue, and a cDNA library was constructed in the Express 1 plasmid vector by Open Biosystems. First strand cDNA synthesis was primed with an 54 bp linker primer containing an oligodT sequence preceded by a synthetic NotI site (first strand primer: 5'-GACTAGTTCAGATCGGACGGCCGCTT25-3'). Following second strand synthesis, cDNAs were made blunt at the end corresponding to the original 5' prime end of mRNA, and cloned directionally into the NotI and EcoRV sites of Express 1. Note that the EcoRV site is typically destroyed in the blunt end cloning, leaving a junction of the form 'xxATC' (where is ATC is the second half of the EcoRV site, and xxx is derived from the cDNA sequence). A map of the Express 1 vector is available at:
http://www.openbiosystems.com/cdna_library_construction_fa_q.php#8 The primary library was transformed and amplified in DH10B (T1 phage resistant) bacteria. Clones available from Open Biosystems:
<http://www.openbiosystems.com/stickleback>"

ORIGIN

Query Match 8.2%; Score 56.2; DB 8; Length 1366;
 Best Local Similarity 31.4%; Pred. No. 0.0057;
 Matches 175; Conservative 0; Mismatches 383; Indels 0; Gaps 0;

13 AATAGAGATTTCAGCAAGTAACGCAAGAAAGGACCTTGAAATTCGTTAAAGAT 72
 1360 AAAAAAAAAANANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1301
 73 GGTATCATCAATCAAGTAACGTAAGTTGATTAATCTCAATACCGGTATCAGTG 132
 1300 NNN 1241
 133 GCTAAGCAACGCGTATTCAGTAGGAGCAAGCGTACCGATCTCAAAATTTTC 192
 1240 AAAAAAAAAANANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1181
 193 TCAGAGAGCAATGGCCCAACAGCTCAAAATGAAAGTCTCAATGCTAGAAAAA 252
 1180 NNN 1121
 253 TCTGAATATATCAATCCCTTAAGATGCTGATGGAACCTAGTCGTTAATGGTTA 312
 1120 AAAAAAAAAANANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1061
 313 TCTCAGCAGAACCACTCTTTCTAATAAATCTTTCGACATCAAGAGATTGAAT 372
 1060 AAAAAAAAAANANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1001
 373 GCAAAATCTTGAATTTCAATTAACAATAATGAATCAAAAGCAAGCCATTAT 432
 1000 NNN 941
 433 GCTAAGTTAATTAAGAAAGAGGAGAGCTAGCTTGAAGAACCTTTACGCT 492
 940 AAAAAAAAAANANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 881
 493 CAAGTTGCTAAAGGTAATGCAAAATTTGACCGATCAATCAATGCAAGTGTG 552
 880 NNN 821
 553 GGTGTTGTAAGGCAAGCA 570
 820 NNN

RESULT 10
 AG430154/c 1400 bp DNA linear GSS 21-DEC-2004
 LOCUS AG430154/c
 DEFINITION Mus musculus molossinus DNA, clone:MSMg01-305A24.TU, genomic survey sequence.

ACCESSION AG430154
 VERSION AG430154.1 GI:48073217
 KEYWORDS GSS.
 SOURCE Mus musculus molossinus (Japanese wild mouse)
 ORGANISM Mus musculus molossinus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
 Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
 Genome Res. 14 (12), 2439-2447 (2004)
 2 (bases 1 to 1400)
 15574823
 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
 Direct Submission
 Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suehiro-chou,Tsuri-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
 Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp). The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan
 phone: 81-298-36-9189, fax: 81-298-36-9199
 e-mail: abe@rtc.riken.jp
 PRIMERS
 Sequencing : TJ
 LIBRARY
 Vector : pBAC3.6
 R.site 1 : EcoRI
 R.site 2 : EcoRI.

COMMENT

FEATURES
 source
 1..1400
 /organism="Mus musculus molossinus"
 /mol_type="genomic DNA"
 /bud_species="molossinus"
 /db_xref="taxon:57486"
 /clone="MSMg01-305A24.TU"
 /sex="male"
 /tissue_type="mixture of kidney and spleen"
 /clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN

Query Match 8.1%; Score 55.6; DB 10; Length 1400;
 Best Local Similarity 29.7%; Pred. No. 0.0081;
 Matches 160; Conservative 0; Mismatches 378; Indels 0; Gaps 0;

1 AAAAAATGCAAAATTAAGATTTCAGCAAGTAACGCAAGAAAGCAAGCTTGAAAT 60
 1175 AAAAAAAAAANANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1116
 61 TCCGTTAAGATGATCAATCAATCAAAAGTAACGATTAAGTTAATCTCAATCAA 120
 1115 AAAAAAAAAANANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1056
 121 GCGGTATCAGTGTCTAAGCAACGCGTATTTTCAGTAGGTAAGAGGCTTAGCGAT 180
 1055 AAAAAAAAAANANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 996
 181 CTCAAAATTTCTCAAGAGCAATGGCCCAACAGCTCAAAAAATGAAGTCTCAAT 240
 995 AAAAAAAAAANANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 936
 241 GCTGAAGAAAAATCTGAATATATCAATCCGTTAAGATGCTGATGGAACCTTAGTC 300
 935 NNN 876


```

QY 361 AAAGAGTGAATGCAGAACTTGGAATTTCATATACATATACATATATGACTCAAAAAC 420
Db 702 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 643
QY 421 GAACCCATTATGCTAAAGTTAATAAAAAGAAAGCAGGCAAGCAGCTAGCTTGAGAA 480
Db 642 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 583
QY 481 CCCATTAGGCTCAAGTGTCTAAAGGTAAATGCAGAAATTGACGACTCATCAAA 538
Db 582 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 525

```

Search completed: March 6, 2006, 22:31:18
 Job time : 3711 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 6, 2006, 22:11:09 ; Search time 351 Seconds
(without alignments)
4278.782 Million cell updates/sec

Title: US-09-360-685C-26

Sequence: 1 aaaaatgcaaaaataagga.....caatcaagcgatcagaag 685

Scoring table: IDENTITY_NIC
Gapop 10.0, Gapext 1.0

Searched: 7218535 seqs, 1096242582 residues

Total number of hits satisfying chosen parameters: 14437070

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries.

Database: Published Applications NA New:*

- 1: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score distribution, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	685	100.0	5925 8 US-10-615-668-4	Sequence 4, Appl1
2	477.2	69.7	3504 12 US-11-052-554A-505	Sequence 505, App
3	48.2	7.0	571 6 US-09-925-065A-20908	Sequence 20908, A
4	46.2	6.7	3143 12 US-11-165-819-9	Sequence 9, Appl1
5	46.2	6.7	3143 12 US-11-165-819-29	Sequence 29, Appl1
6	45	6.6	45 8 US-10-615-668-22	Sequence 22, Appl1
7	43.4	6.3	19233 8 US-10-240-708-45	Sequence 45, Appl1
8	43	6.3	648 6 US-09-925-065A-64219	Sequence 64219, A
9	42.2	6.2	533 6 US-09-925-065A-127945	Sequence 127945, A
10	42.2	6.2	633 6 US-09-925-065A-140562	Sequence 140562, A
11	41.2	6.0	524 6 US-09-925-065A-575490	Sequence 575490, A
12	40.6	5.9	10467 8 US-10-240-708-1	Sequence 1, Appl1
13	39.8	5.8	598 6 US-09-925-065A-410917	Sequence 410917, A
14	39.8	5.8	670 6 US-09-925-065A-474252	Sequence 474252, A
15	39.8	5.8	670 6 US-09-925-065A-474253	Sequence 474253, A
16	39.6	5.8	102 8 US-10-615-668-11	Sequence 11, Appl1
17	39.6	5.8	3221 8 US-10-793-626-4163	Sequence 4163, Ap
18	39.6	5.8	3423 8 US-10-793-626-3355	Sequence 3355, Ap
19	39.2	5.7	539 6 US-09-925-065A-237166	Sequence 237166, A
20	39.2	5.7	605 6 US-09-925-065A-524453	Sequence 524453, A

21	39.2	5.7	1488 6 US-09-925-065A-51174	Sequence 51174, A
22	39.2	5.7	1488 6 US-09-925-065A-51175	Sequence 51175, A
23	39.2	5.7	1488 6 US-09-925-065A-51176	Sequence 51176, A
24	39	5.7	2407 8 US-10-485-517-42	Sequence 42, Appl1
25	39	5.7	9347 8 US-10-240-708-35	Sequence 35, Appl1
26	38.8	5.7	605 6 US-09-925-065A-524454	Sequence 524454, A
27	38.6	5.6	45 8 US-10-615-668-18	Sequence 18, Appl1
28	38.6	5.6	487 6 US-09-925-065A-304585	Sequence 304585, A
29	38.6	5.6	539 6 US-09-925-065A-328934	Sequence 328934, A
30	38.6	5.6	539 6 US-09-925-065A-328935	Sequence 328935, A
31	38.6	5.6	568 6 US-09-925-065A-107617	Sequence 107617, A
32	38.6	5.6	598 6 US-09-925-065A-410918	Sequence 410918, A
33	38.6	5.6	179597 12 US-11-121-086-91	Sequence 91, Appl1
34	38.4	5.6	622 6 US-09-925-065A-873655	Sequence 873655, A
35	38.4	5.6	644 6 US-09-925-065A-418161	Sequence 418161, A
36	38.4	5.6	1290 6 US-09-925-065A-705891	Sequence 705891, A
37	38.4	5.6	10144 8 US-10-240-708-93	Sequence 93, Appl1
38	38.4	5.6	1131 8 US-10-240-708-27	Sequence 27, Appl1
39	38.4	5.5	842 8 US-10-750-185-42851	Sequence 42851, A
40	38	5.5	842 8 US-10-750-185-42851	Sequence 42851, A
41	38	5.5	954 8 US-10-750-185-60688	Sequence 60688, A
42	38	5.5	954 8 US-10-750-185-60688	Sequence 60688, A
43	38	5.5	139054 12 US-11-121-086-96	Sequence 96, Appl1
44	38	5.5	631 6 US-09-925-065A-215960	Sequence 215960, A
45	37.8	5.5		

ALIGNMENTS

RESULT 1				
US-10-615-668-4				
Sequence 4, Application US/10615668				
Publication No. US20050276819A1				
GENERAL INFORMATION:				
APPLICANT: Bugnoli, Massimo				
APPLICANT: Bugnoli, Massimo				
APPLICANT: Telford, John				
APPLICANT: Macchia, Giovanni				
TITLE OF INVENTION: Helicobacter Pylori CAI Antigen Polynucleotides				
FILE REFERENCE: CHIR0337				
CURRENT APPLICATION NUMBER: US/10/615, 668				
CURRENT FILING DATE: 2003-07-08				
PRIOR APPLICATION NUMBER: 08/471,491				
PRIOR FILING DATE: 1995-06-06				
PRIOR APPLICATION NUMBER: 08/256, 848				
PRIOR FILING DATE: 1994-10-21				
PRIOR APPLICATION NUMBER: 09/410, 835				
PRIOR FILING DATE: 1999-10-01				
NUMBER OF SEQ ID NOS: 24				
SOFTWARE: PatentIn version 3.2				
SEQ ID NO 4				
LENGTH: 5925				
TYPE: DNA				
ORGANISM: Helicobacter pylori				
US-10-615-668-4				
Query Match				
Best Local Similarity 100.0%; Pred. No. 1.1e-147; Length 5925;				
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1	AAAAATGGCAAAATAGATTTCGCAAGTACGCAAGCAAAAGCAGCTTGAAT	60	
DB	2782	AAAAATGGCAAAATAGATTTCGCAAGTACGCAAGCAAAAGCAGCTTGAAT	2841	
QY	61	TCCGTTAAAGATGATCATCATATAAAGTAAAGTAAAGTTGATTAATCAATCA	120	
DB	2842	TCCGTTAAAGATGATCATCATATAAAGTAAAGTAAAGTTGATTAATCAATCA	2901	
QY	121	GGGATATAGTGGTAAAGCAAGGATTTCACTAGGATTAAGCAAGCTTACCGAT	180	
DB	2902	GGGATATAGTGGTAAAGCAAGGATTTCACTAGGATTAAGCAAGCTTACCGAT	2961	

QY 181 CTCGAAATTTCTCGAAGGAGCAATGGCCCAAGACTCAAAAAATGAAAGTCTCAAT 240
DB 2962 CTCGAAATTTCTCGAAGGAGCAATGGCCCAAGACTCAAAAAATGAAAGTCTCAAT 240
QY 241 GCTGAAAAAATCTGAAATATATCAATCCGTTAAGAAATGATGAAACCTTAATC 300
DB 3022 GCTGAAAAAATCTGAAATATATCAATCCGTTAAGAAATGATGAAACCTTAATC 300
QY 301 GGTAAATGGTTATCTCAAGCAAGAACCAACTCTTTCTAAAAACTTTCCGACATCAAG 360
DB 3082 GGTAAATGGTTATCTCAAGCAAGAACCAACTCTTTCTAAAAACTTTCCGACATCAAG 360
QY 361 AAGAGTTGAATGCAAACTTGAAATTTCAATTAACAATTAATGAGCTCAAAAAC 420
DB 3142 AAGAGTTGAATGCAAACTTGAAATTTCAATTAACAATTAATGAGCTCAAAAAC 420
QY 421 GAACCATTTATGCTAAAGTTAATAAAGAAAGGAGGCAAGAGCTAGCTTGAAGA 480
DB 3202 GAACCATTTATGCTAAAGTTAATAAAGAAAGGAGGCAAGAGCTAGCTTGAAGA 480
QY 481 CCCATTTACGCTCAAGTTGCTAAAAAGTTAATAAAGAAAGGAGGCAAGAGCTAGCTTGAAGA 540
DB 3262 CCCATTTACGCTCAAGTTGCTAAAAAGTTAATAAAGAAAGGAGGCAAGAGCTAGCTTGAAGA 540
QY 541 GCAAGTGGTTGGTGTGTGAGGCAAGAGGAGGCTTCCCTTTGAAAAGGATGATAA 600
DB 3322 GCAAGTGGTTGGTGTGTGAGGCAAGAGGAGGCTTCCCTTTGAAAAGGATGATAA 600
QY 601 GTTATGATCTCAAGTAAAGTTGAGGCTTTCAAGAAATCAAGAAATGAGCTCAAGAAATGAG 660
DB 3382 GTTATGATCTCAAGTAAAGTTGAGGCTTTCAAGAAATCAAGAAATGAGCTCAAGAAATGAG 660
QY 661 AATCTCAATCAAGCGCTTCAAG 685
DB 3442 AATCTCAATCAAGCGCTTCAAG 685

RESULT 2
US-11-052-554A-505
/ Sequence 505, Application US/11052554A
/ Publication No. US2005028866A1
/ GENERAL INFORMATION:
/ APPLICANT: Sachdeva, et al.
/ TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
/ FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
/ FILE REFERENCE: 30853/40359A
/ CURRENT APPLICATION NUMBER: US/11/052,554A
/ PRIOR FILING DATE: 2005-02-07
/ PRIOR APPLICATION NUMBER: US 60/589,227
/ PRIOR FILING DATE: 2004-07-20
/ PRIOR APPLICATION NUMBER: IN 173/DEL/2004
/ PRIOR FILING DATE: 2004-02-06
/ NUMBER OF SEQ ID NOS: 763
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 505
/ LENGTH: 3504
/ TYPE: DNA
/ ORGANISM: Helicobacter pylori J99
US-11-052-554A-505

Query Match 69.7%; Score 477.2; DB 12; Length 3504;
Best Local Similarity 86.9%; Pred. No. 5.3e-100;
Matches 558; Conservative 0; Mismatches 33; Indels 51; Gaps 1;
QY 1 AAAAAATGCAAAATTAAGATTTTCGCAAGTTACGCAAGCAAAAAAGCACTTGAAT 60
DB 2260 AAAAAATGCAAAATTAAGATTTTCGCAAGTTACGCAAGCAAAAAAGCACTTGAAT 60
QY 61 TCGTAAAGATGATCATCAATCAAAAGTAAAGGATTAAGTTGATTAATCTCAATCA 120
DB 2320 TCGTAAAGATGATCATCAATCAAAAGTAAAGGATTAAGTTGATTAATCTCAATCA 120

QY 121 GCGTATCAGTGGCTAAAGCAAGGATGATTTGAGTAAAGGTAAGCAAGGTTAGCCGAT 180
DB 2380 GCGTATCAGTGGCTAAAGCAAGGATGATTTGAGTAAAGGTAAGCAAGGTTAGCCGAT 180
QY 181 CTCGAAATTTCTCGAAGGAGCAATGGCCCAAGACTCAAAAAATGAAAGTCTCAAT 240
DB 2440 CTCGAAATTTCTCGAAGGAGCAATGGCCCAAGACTCAAAAAATGAAAGTCTCAAT 240
QY 241 GCTGAAAAAATCTGAAATATATCAATCCGTTAAGAAATGATGAAACCTTAATC 300
DB 2500 ACTGAAAAAATCTGAAATATATCAATCCGTTAAGAAATGATGAAACCTTAATC 300
QY 301 GGTAAATGGTTATCTCAAGCAAGAACCAACTCTTTCTAAAAACTTTCCGACATCAAG 360
DB 2560 GGTAAATGGTTATCTCAAGCAAGAACCAACTCTTTCTAAAAACTTTCCGACATCAAG 360
QY 361 AAGAGTTGAATGCAAACTTGAAATTTCAATTAACAATTAATGAGCTCAAAAAC 420
DB 2620 AAGAGTTGAATGCAAACTTGAAATTTCAATTAACAATTAATGAGCTCAAAAAC 420
QY 421 GAACCATTTATGCTAAAGTTAATAAAGAAAGGAGGCAAGAGCTAGCTTGAAGA 480
DB 2680 GAACCATTTATGCTAAAGTTAATAAAGAAAGGAGGCAAGAGCTAGCTTGAAGA 480
QY 481 CCCATTTACGCTCAAGTTGCTAAAAAGTTAATAAAGAAAGGAGGCAAGAGCTAGCTTGAAGA 540
DB 2688 CCCATTTACGCTCAAGTTGCTAAAAAGTTAATAAAGAAAGGAGGCAAGAGCTAGCTTGAAGA 540
QY 541 GCAAGTGGTTGGTGTGTGAGGCAAGAGGAGGCTTCCCTTTGAAAAGGATGATAA 600
DB 2748 GCAAGTGGTTGGTGTGTGAGGCAAGAGGAGGCTTCCCTTTGAAAAGGATGATAA 600
QY 601 GTTATGATCTCAAGTAAAGTTGAGGCTTTCAAGAAATCAAGAAATGAGCTCAAGAAATGAG 660
DB 2808 GTTATGATCTCAAGTAAAGTTGAGGCTTTCAAGAAATCAAGAAATGAGCTCAAGAAATGAG 660

RESULT 3
US-09-925-065A-20908
/ Sequence 20908, Application US/09925065A
/ Publication No. US20040181048A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single
/ FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
/ FILE REFERENCE: 10827.135
/ CURRENT APPLICATION NUMBER: US/09/925,065A
/ PRIOR FILING DATE: 2001-08-08
/ PRIOR APPLICATION NUMBER: US 60/243,096
/ PRIOR FILING DATE: 2000-10-24
/ PRIOR APPLICATION NUMBER: US 60/252,147
/ PRIOR FILING DATE: 2000-11-20
/ PRIOR APPLICATION NUMBER: US 60/250,092
/ PRIOR FILING DATE: 2000-11-30
/ PRIOR APPLICATION NUMBER: US 60/261,766
/ PRIOR FILING DATE: 2001-01-16
/ PRIOR APPLICATION NUMBER: US 60/289,846
/ PRIOR FILING DATE: 2001-05-09
/ NUMBER OF SEQ ID NOS: 957086
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 20908
/ LENGTH: 571
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-925-065A-20908

Query Match 7.0%; Score 48.2; DB 6; Length 571;
Best Local Similarity 50.7%; Pred. No. 0.11; Indels 0; Gaps 0;
Matches 116; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 221 AAAAAATGAAAGCTCAATGCTTCAAAAAATGCAATATATCAATCCGTTAAGATG 280
DB 24 AAAAAATGAAAGCTCAATGCTTCAAAAAATGCAATATATCAATCCGTTAAGATG 280

RESULT 5
US-11-165-819-29
? Sequence 29, Application US/11165819
? Publication No. US20060019314A1
? GENERAL INFORMATION:
? APPLICANT: Ma, Hui
? APPLICANT: Dillin, Andrew

```

Sequence 22, Application US/10615668
Publication No. US20050276619A1
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Helicobacter Pylori CAI Antigen Polynucleotides
FILE REFERENCE: CHIR0337
CURRENT APPLICATION NUMBER: US/10/615,668
CURRENT FILING DATE: 2003-07-08
PRIOR APPLICATION NUMBER: 08/471,431
PRIOR FILING DATE: 1995-06-06
PRIOR APPLICATION NUMBER: 08/256,848
PRIOR FILING DATE: 1994-10-21
PRIOR APPLICATION NUMBER: 09/410,835
PRIOR FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.2
SEQ ID NO 22
LENGTH: 45
TYPE: DNA
ORGANISM: Helicobacter pylori
US-10-615-668-22

```



```
Db 340 ATACGAGAGCTGAAAAAATGATGATCAAAAAACAAGCAAGAAAGTGA 281
Qy 239 ATCTAGAAAAAATCTGAATATATCAATCCGTAGAAAGGTGATGAAACCTAG 298
Db 280 GAATATATAATTTTGTGACATATAGAAATGAAATTTGATATGATGATTAACCA 221
Qy 299 TCGGTAATGGTTATCTCAAGCAGACCAACTCTTTTAAAACTTTTGGACATCA 358
Db 220 TATGGAGTACTTTCATTAAGATGAGGCAAAAAATTAACCAAAATTTGTGGAAGTTG 161
Qy 359 AGAAAGGTGATGCAAACTTGAATTTCAATTAACAATTAATGACTCAAAA 418
Db 160 TTGAATTTGATGAAAAATCATGCTCATCAAGACGATATCTTAACATATGATCTC 101
Qy 419 ACGAACCCATTATGCTAAGTTAATAAAAA 453
Db 100 AGAATGTTTCAAAAATTAAGATGAAAAAGATA 66
```

RESULT 9

```
US-09-925-065A-127945/c
; Sequence 127945, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243, 096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252, 147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250, 092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261, 766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289, 846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 127945
; LENGTH: 533
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-127945
```

Query Match 6.2%; Score 42.2; DB 6; Length 533;

Best Local Similarity 54.2%; Pred. No. 2.5; Matches 83; Conservative 1; Mismatches 69; Indels 0; Gaps 0;

```
Qy 305 ATGGGTTATCTCAAGCAGACCAACTCTTTTAAAACTTTTGGACATCAAGAAAG 364
Db 260 ATTAGGTGTTAAAGGTAATGTCATATTAAGTAATTTTGAACCTAAGAAACG 201
Qy 365 AGTTGATGCAAACTTGAATTTCAATTAACAATTAATGACTCAAAAAGAAC 424
Db 200 TTAGAAATCATGTATGAAATTTTTCACAGATAAAAAGGTGATTAATATCTCTC 141
Qy 425 CCATTATGCTAAGTTAATAAAAAAGAACG 457
Db 140 CAAATAATTCACAAAATTAATTAAGAAATTAACG 108
```

RESULT 10

```
US-09-925-065A-140562/c
; Sequence 140562, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
```

```
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243, 096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252, 147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250, 092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261, 766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289, 846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140562
; LENGTH: 633
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-140562
```

Query Match 6.2%; Score 42.2; DB 6; Length 633;

Best Local Similarity 55.1%; Pred. No. 2.6; Matches 103; Conservative 0; Mismatches 83; Indels 1; Gaps 1;

```
Qy 302 GTAATGGTTATCTCAAGCAGACCAACTCTTTTAAAACTTTTGGACATCAAGA 361
Db 475 GCAAAAAGTTATCTTATGCAAAATGCAAAATGCAACAAAGACTTATCAATTAAGCA 416
Qy 362 AAGAGTTGATGCAAACTTGAATTTCAATTAACAATTAATGACTCAAAAAGC 421
Db 415 AATAGTTTAACTTAAATTAATTAACAAGACAAATTAAC-ATTGTGTAATGATTAAG 357
Qy 422 AACCCATTATGCTAAGTTAATTAATTAAGAAAGCAGGAGAGCTAGCTTGAAGAC 481
Db 356 AATCATTCATCGAAGAGTACATTAATTAATTAATTAATGACCTATCATATGAAAC 297
Qy 482 CCATTTA 488
Db 296 ATTAATA 290
```

RESULT 11

```
US-09-925-065A-575490
; Sequence 575490, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
```

```
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243, 096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252, 147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250, 092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261, 766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289, 846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 575490
; LENGTH: 524
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-575490
```

Query Match

6.0%; Score 41.2; DB 6; Length 524;

Tue Mar 7 12:58:26 2006

us-09-360-685c-26.rnpbn

Page 6

	Best Local Similarity	46.2%	Pred. No. 4.2,
	Matches 168;	Conservative 1;	Mismatches 194; Indels 1; Gaps 1
QY	194	CAAGAGCAATGTGGCCCAACAGCTCCAAAATGAAAGTCTGATGCTGAAAAAT	253
Db	36	CAAAACAAACAAAAAGTGAACAAATTTTGTGAAATGATTAAGAAAGCAAAAAGAAA	95
QY	254	CTGAAATATATCAATCCGTTAGAAATGTGTGAAATGGAACCCATGTCGTAATGGCTAT	313
Db	96	AAAACTCAACATATCTGAAAAAAGTAAATTAATTAATTTAGCATTTGACGGCATTAACGAAA	155
QY	314	CTCAGACAGACCCGCAACTCTTCTGAAAACTTTTCGACATCAAGAAAGATTGATG	373
Db	156	TCTTTATGAGCAATTAAGAAATATGTGATTAATTTGTGATTA-ATTAATTTTAAACATTAA	214
QY	374	CAAAATTTGGAAATTTCAATTAACATTAACATATATGATCTCAAAAAACGAAACCATTTAG	433
Db	215	TATATATGACCAATTTACTGAAAGAAATTCGATTAATTTCAAGAAATATGATTAATCAACTA	274
QY	434	CTTAAGTTTATTAATAAAGAAAGCAAGGGCAAGCACTGAGCTTGTGAAGAACCCATTAGAGTC	493
Db	275	CTACTGTATCTACTCAATTAACAGTTTATGCGCAACAGCTTTAAAAATCTATGTCTCAAGAAAGTT	334
QY	494	AAGTTGCTTAATAAGGTAAATGTGCAAAAAATTTGACCGATCATCAATTAATACCAAGTCGTTGG	553
Db	335	CAGGTCAACAGAGCTTCAATCCCAAAAAAATTCATTAATGATCTCTTAAGAAATTAATTTTA	394
QY	554	GTGT 557	
Db	395	GTTT 398	

```

RESULT 12
US-10-240-708-1/c
Sequence 1 Application US/10240708
Publication No US20050282157A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIESENROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/240,708
PRIORITY FILING DATE: 2002-10-03
PRIORITY APPLICATION NUMBER: PCT/EP01/03971
PRIORITY FILING DATE: 2001-04-06
PRIORITY APPLICATION NUMBER: DE 10019058.8
PRIORITY FILING DATE: 2000-04-06
PRIORITY APPLICATION NUMBER: DE 10019173.8
PRIORITY FILING DATE: 2000-04-07
PRIORITY APPLICATION NUMBER: DE 10032529.7
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: DE 10043826.1
PRIORITY FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 1
LENGTH: 10467
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
US-10-240-708-1
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

```

```

Query Match      5.94; Score 40.6; DB 8; Length 10467;
Best local similarity 50.84; P-adj. No. 15;
Matches 97; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY      345 CTTTGGGACATCGAAGAGCTTGATGCGAAACTGTGAAATTCATATACATACAA 404
      |||||
      10178 CTTTACAAATTTTAAAAATTTTAAAAATTTTCAATTCACAAAACATTTACC 10118
QY      405 TATGCACTCAAAAGAACCCATTTCGTAAAGTTAATTAATTAAGAAAGCAAGGCGACG 464

```

[illegible]

```

RESULT 13
/ US-09-925-065A-410917
/ Sequence 410917, Application US /09925065A
/ Publication No. US20040181048A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single
/ FILE REFERENCE: 108827.135
/ CURRENT APPLICATION NUMBER: US/09/925,065A
/ CURRENT FILING DATE: 2001-08-08
/ PRIOR APPLICATION NUMBER: US 60/243,096
/ PRIOR FILING DATE: 2000-10-24
/ PRIOR APPLICATION NUMBER: US 60/252,147
/ PRIOR FILING DATE: 2000-11-20
/ PRIOR APPLICATION NUMBER: US 60/250,092
/ PRIOR FILING DATE: 2000-11-30
/ PRIOR APPLICATION NUMBER: US 60/261,766
/ PRIOR FILING DATE: 2001-01-16
/ PRIOR APPLICATION NUMBER: US 60/289,846
/ PRIOR FILING DATE: 2001-05-09
/ NUMBER OF SEQ ID NOS: 957086
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 410917
/ LENGTH: 598
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-09-925-065A-410917

```

Query Match	Best Local Similarity	5.84; Score 39.8; DB 6; Length 598;
Matches 124; Conservative	1; Mismatches 123; Indels 1; Gaps 1;	
Qy	195	AAAGAGCAATTGGCCCAACAGCTCAAAAAATGAAAGTCTCAATGTACGAAAAAAATC 254
Db	114	AATGAGAAAAAGGAGAAAAATGCAAAAAAAATTTGAATATAAAGTAAATTCAGAAA 173
Qy	255	TGAATATATCATCCGTTAAGATGCTGATGGAACCCGTGCGTAAAGGTTATTC 314
Db	174	TGACAAAGTATTTTAAAAAATATTAAAGTATTAACATTTCCATTATTTACAACT 233
Qy	315	TCAAGCAGAACCAACTCTTTCTTAAAAACTTTTCGGACATCAAAAGAGTTGAATC 374
Db	234	TTAATGAGGTAGTACTTTTTTTGAARAAAGAACTTCGAAATGACAT- TGAAGATGG 292
Qy	375	AAAACTGGAAATTTCAATTAACATTAACATTAATGAGCTGAAAACGAAACCATTTATGC 434
Db	293	AAATATCAAAAAATTTATTAATTAACATTAATGAAAGAAATTTTAAAAACAAATTAACCATTAATC 352
Qy	435	TAAAGTTAA 443
Db	353	CAAAAAATTA 361

RESULT 14
 US-09-925-065A-474252/C
 ; Sequence 474252, Application US/09925065A
 ; Publication No. US20040181048A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, David G.
 ; TITLE OF INVENTION: Identification and Mapping of Single

! TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
! FILE REFERENCE: 108827.135
! CURRENT APPLICATION NUMBER: US/09/925,065A
! CURRENT FILING DATE: 2001-08-08
! PRIOR APPLICATION NUMBER: US 60/243,096
! PRIOR FILING DATE: 2000-10-24
! PRIOR APPLICATION NUMBER: US 60/252,147
! PRIOR FILING DATE: 2000-11-20
! PRIOR APPLICATION NUMBER: US 60/250,092
! PRIOR FILING DATE: 2000-11-30
! PRIOR APPLICATION NUMBER: US 60/261,766
! PRIOR FILING DATE: 2001-01-16
! PRIOR APPLICATION NUMBER: US 60/289,846
! PRIOR FILING DATE: 2001-05-09
! NUMBER OF SEQ ID NOS: 957086
! SOFTWARE: FastSeq for Windows Version 4.0
! SEQ ID NO 474252
! LENGTH: 670
! TYPE: DNA
! ORGANISM: Homo sapiens
US-09-925-065A-474252

Query Match 5.8%; Score 39.8; DB 6; Length 670;
Best Local Similarity 54.4%; Pred. No. 9.5;
Matches 80; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 375 AAACCTGGAATTTCAATACATATATGACCTCAAAAGCAACCCATTATGC 434
DB 584 AAGCAGAGATATAGAAATGCGATTTGAAAAAGAACCAAGAAATCTCCATATATGA 525
QY 435 TAAAGTTAATTAAGAAAGCAGGCGACGCTTGAAGAACCCATTACGCTCA 494
DB 524 AAAATATATTAATTTGAATCCAACTTAACAGATACGTTAAAAATTAACATTATAGAAAT 465
QY 495 AGTTGCTAAAGGTTAATGCAAAAT 521
DB 464 AGCTGATTAAGAAAGACATTTTAAAAAT 438

RESULT 15
US-09-925-065A-474253/c
! Sequence 474253, Application US/09925065A
! Publication No. US20040181048A1
! GENERAL INFORMATION:
! APPLICANT: Wang, David G.
! TITLE OF INVENTION: Identification and Mapping of Single
! FILE REFERENCE: 108827.135
! Nucleotide Polymorphisms in the Human Genome
! CURRENT APPLICATION NUMBER: US/09/925,065A
! CURRENT FILING DATE: 2001-08-08
! PRIOR APPLICATION NUMBER: US 60/243,096
! PRIOR FILING DATE: 2000-10-24
! PRIOR APPLICATION NUMBER: US 60/252,147
! PRIOR FILING DATE: 2000-11-20
! PRIOR APPLICATION NUMBER: US 60/250,092
! PRIOR FILING DATE: 2000-11-30
! PRIOR APPLICATION NUMBER: US 60/261,766
! PRIOR FILING DATE: 2001-01-16
! PRIOR APPLICATION NUMBER: US 60/289,846
! PRIOR FILING DATE: 2001-05-09
! NUMBER OF SEQ ID NOS: 957086
! SOFTWARE: FastSeq for Windows Version 4.0
! SEQ ID NO 474253
! LENGTH: 670
! TYPE: DNA
! ORGANISM: Homo sapiens
US-09-925-065A-474253

Query Match 5.8%; Score 39.8; DB 6; Length 670;
Best Local Similarity 54.4%; Pred. No. 9.5;
Matches 80; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 375 AAACCTGGAATTTCAATATACATATATGACCTCAAAAGCAACCCATTATGC 434

DB 584 AAGCAGAGATATAGAAATGCGATTTGAAAAAGAACCAAGAAATCTCCATATATGA 525
QY 435 TAAAGTTAATTAAGAAAGCAGGCGACGCTTGAAGAACCCATTACGCTCA 494
DB 524 AAAATATATTAATTTGAATCCAACTTAACAGATACGTTAAAAATTAACATTATAGAAAT 465
QY 495 AGTTGCTAAAGGTTAATGCAAAAT 521
DB 464 AGCTGATTAAGAAAGACATTTTAAAAAT 438

Search completed: March 6, 2006, 22:31:20
Job time : 352 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2006, 22:06:00 ; Search time 782 Seconds

(without alignments)
7243.637 Million cell updates/sec

Title: US-09-360-685C-26

Perfect score: 685
Sequence: 1 aaaaatgcgaataaaga.....caatcaagcgatcagaag 685

Scoring table: IDENTITY_NIC
Gapop 10.0, Gapept 1.0

Searched: 9793542 seqs, 413468905 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: Published Applications NA Main:

1: /cgn2_6/prodata/1/pubphn/US07_PUBCOMB.seq.*
2: /cgn2_6/prodata/1/pubphn/US08_PUBCOMB.seq.*
3: /cgn2_6/prodata/1/pubphn/US09_PUBCOMB.seq.*
4: /cgn2_6/prodata/1/pubphn/US10_PUBCOMB.seq.*
5: /cgn2_6/prodata/1/pubphn/US10A_PUBCOMB.seq.*
6: /cgn2_6/prodata/1/pubphn/US10B_PUBCOMB.seq.*
7: /cgn2_6/prodata/1/pubphn/US10C_PUBCOMB.seq.*
8: /cgn2_6/prodata/1/pubphn/US10D_PUBCOMB.seq.*
9: /cgn2_6/prodata/1/pubphn/US10E_PUBCOMB.seq.*
10: /cgn2_6/prodata/1/pubphn/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	685	100.0	4149	3	US-09-402-100-3 Sequence 3, Appl1
2	685	100.0	5925	3	US-09-921-157-4 Sequence 4, Appl1
3	477.2	69.7	3504	7	US-10-335-977-3871 Sequence 3871, Ap
4	477.2	69.7	3534	7	US-10-335-977-3872 Sequence 3872, Ap
5	477.2	69.7	3549	7	US-10-335-977-3873 Sequence 3873, Ap
6	50.6	7.4	3673778	6	US-10-312-841-1 Sequence 1, Appl1
7	48.2	7.0	571	4	US-09-925-065A-20908 Sequence 20908, A
8	48	7.0	12138	6	US-10-311-455-1916 Sequence 1916, Ap
9	48	7.0	12138	6	US-10-240-453-210 Sequence 210, Appl1
10	47.8	7.0	3673778	6	US-10-312-841-2 Sequence 1931, Ap
11	47	6.9	14006	6	US-10-311-455-1931 Sequence 1931, Ap
12	46.8	6.8	5542	6	US-10-311-455-1933 Sequence 1933, Ap
13	46.8	6.8	9964	6	US-10-311-455-71 Sequence 71, Appl1
14	46.4	6.8	921	8	US-10-425-115-38710 Sequence 38710, A
15	46.4	6.8	5127	5	US-10-239-676-132 Sequence 132, Appl1
16	46	6.7	7969	7	US-10-221-714A-401 Sequence 401, Appl1
17	46	6.7	7969	7	US-10-433-793-67 Sequence 67, Appl1
18	45.6	6.7	6182	6	US-10-311-455-1988 Sequence 1988, Ap
19	45.4	6.6	14362	7	US-10-333-793-78 Sequence 78, Appl1
20	45.4	6.6	14708	5	US-10-339-676-222 Sequence 222, Appl1
21	45.4	6.6	14708	5	US-10-311-455-2218 Sequence 2218, Ap
22	45.4	6.6	14708	6	US-10-240-453-324 Sequence 324, Appl1
23	45.4	6.6	14708	7	US-10-221-714A-500 Sequence 500, Appl1

C	24	45.4	6.6	14708	7	US-10-240-589C-134	Sequence 134, Appl1
C	25	45.2	6.6	8222	9	US-10-486-319A-43	Sequence 43, Appl1
C	26	45.2	6.6	8222	9	US-10-486-319A-55	Sequence 65, Appl1
C	27	45.2	6.6	5814	6	US-10-311-455-1534	Sequence 1534, Ap
C	28	45	6.6	9666	6	US-10-311-455-1058	Sequence 1058, Ap
C	29	45	6.6	9666	6	US-10-240-485-84	Sequence 84, Appl1
C	30	45	6.6	11812	5	US-10-239-676-210	Sequence 210, Appl1
C	31	45	6.6	11812	6	US-10-311-455-2092	Sequence 2092, Ap
C	32	45	6.6	11812	6	US-10-240-453-306	Sequence 306, Appl1
C	33	45	6.6	11812	7	US-10-221-714A-468	Sequence 468, Appl1
C	34	45	6.6	19787	7	US-10-311-455-1423	Sequence 1423, Ap
C	35	44.8	6.5	1332	2	US-08-325-278-5	Sequence 5, Appl1
C	36	44.8	6.5	1662	8	US-10-474-792-671	Sequence 671, Appl1
C	37	44.8	6.5	2543	5	US-10-027-632-102481	Sequence 102481, Appl1
C	38	44.8	6.5	2543	5	US-10-027-632-102482	Sequence 102482, Appl1
C	39	44.8	6.5	2543	6	US-10-027-632-102481	Sequence 102481, Appl1
C	40	44.8	6.5	2543	6	US-10-027-632-102482	Sequence 102482, Appl1
C	41	44.8	6.5	3885	8	US-10-377-636-3	Sequence 3, Appl1
C	42	44.6	6.5	113515	6	US-10-311-455-2147	Sequence 2147, Appl1
C	43	44.2	6.5	6171	6	US-10-311-455-761	Sequence 761, Appl1
C	44	44.2	6.5	7369	8	US-10-473-126-277	Sequence 277, Appl1
C	45	44.2	6.5	7369	9	US-10-486-319A-56	Sequence 56, Appl1

ALIGNMENTS

```
RESULT 1
US-09-402-100-3
; Sequence 3, Application US/09402100
; Patent No. US20010019834A1
; GENERAL INFORMATION:
; APPLICANT: Daewoong Pharmaceutical Co, LTD
; APPLICANT: Kim, Byung-O
; APPLICANT: Shin, Sung-Seup
; APPLICANT: Yu, Young-Hyo
; APPLICANT: Park, Myung-Hwan
; APPLICANT: Choi, Deok-Oon
; TITLE OF INVENTION: Recombinant Microorganisms Expressing Antigenic Proteins of Heli.
; FILE REFERENCE: 0136/06140
; CURRENT APPLICATION NUMBER: US/09/402,100
; CURRENT FILING DATE: 1999-09-27
; EARLIER APPLICATION NUMBER: KR 97-11950
; EARLIER FILING DATE: 1997-03-31
; EARLIER APPLICATION NUMBER: KR 97-11951
; EARLIER FILING DATE: 1997-03-31
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 4149
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (..)..
; OTHER INFORMATION: Recombinant DNA
US-09-402-100-3
Query Match 100.0%; Score 685; DB 3; Length 4149;
Best Local Similarity 100.0%; Pred. No. 3.1e-153;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAAGCAAAAATGAGATTTCGACGAGTAACGCAAGCAAAAAGGACCTGAAT 60
DB AAAAAAGCAAAAATGAGATTTCGACGAGTAACGCAAGCAAAAAGGACCTGAAT 2307
QY 61 TCCGTTAAAGATGATCATCATCAAAAGGTAACGATTAAGTGAATCTCAATCA 120
DB TCCGTTAAAGATGATCATCATCAAAAGGTAACGATTAAGTGAATCTCAATCA 2367
QY 121 GCGGATACGATGCGGTAACGACGGGATTTTCAGTAGGTAAGCAAGCGTAA 180
```

Tue Mar 7 12:58:26 2006

us-09-360-685c-26.rtf

Page 2

Db 2368 GCGGATATCATGGCTTAAAGCAACGGGTGATTTCTAGTAGGGTATGAGCAACGTTAGCCGAT 2427

Qy 181 CTCAAAAATTTCTCAAAGAGCAATTTGGCCCAACAGCTCAAAAAATGAAAGTCTCAAT 240

Db 2428 CTCAAAAATTTCTCAAAGAGCAATTTGGCCCAACAGCTCAAAAAATGAAAGTCTCAAT 2487

Qy 241 GCTGAAAAAAATCTCGAAATTTATCAATCGGTTAGAAATGGTGTGATGSAACCTTAGTC 300

Db 2488 GCTGAAAAAAATCTCGAAATTTATCAATCGGTTAGAAATGGTGTGATGSAACCTTAGTC 2547

Qy 301 GGTAATGGGTATCTCAAGCAAGCAACAATCTTTCTAAAAAATTCTTGGCATCAAG 360

Db 2548 GGTAAATGGGTATCTCAAGCAAGCAACAATCTTTCTAAAAAATTCTTGGCATCAAG 2607

Qy 361 AAAGATGTAAGCAAACTTGGAAATTTCAATACATTAACATATGATGACTCAAAAAC 420

Db 2608 AAAGATGTAAGCAAACTTGGAAATTTCAATACATTAACATATGATGACTCAAAAAC 2667

Qy 421 GAACCCCTTTATGCTAAAGTTATATAAAAAAGACAGGGCAACAGCTAGGCTTGAAGA 480

Db 2668 GAACCCCTTTATGCTAAAGTTATATAAAAAAGACAGGGCAACAGCTAGGCTTGAAGA 2727

Qy 481 GCCCTTTAAGGCTCAAGTGTCTTAAAAAGATAATGCAAAAATTGACCGACTCATCAATA 540

Db 2728 CCCCTTTAAGGCTCAAGTGTCTTAAAAAGATAATGCAAAAATTGACCGACTCATCAATA 2787

Qy 541 GCACAGTGTGGGTGTGTGAGGCAAGCAGGGGCTTCCCTTGAAAAAGCATGTATAA 600

Db 2788 GCACAGTGTGGGTGTGTGAGGCAAGCAGGGGCTTCCCTTGAAAAAGCATGTATAA 2847

Qy 601 GTTGATGATCTCAAGTAAGTAGGGCTTTCAAGATCAAAATTGGCTCGAATAATGAC 660

Db 2848 GTTGATGATCTCAAGTAAGTAGGGCTTTCAAGATCAAAATTGGCTCGAATAATGAC 2907

Qy 661 AATTCATCAAGGGGTATCAAG 685

Db 2908 AATTCATCAAGGGGTATCAAG 2932

```

RESULT 2
US-09-921-157-4
Sequence 4, Application US/09921157
Publication No. US2004004853A1
GENERAL INFORMATION:
APPLICANT: Covacci, Antonio
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Helicobacter Pylori Cytotoxin Proteins Useful For
FILE REFERENCE: CHIR0315
CURRENT APPLICATION NUMBER: US/09/921,157
PRIORITY FILING DATE: 2001-08-02
PRIORITY APPLICATION NUMBER: 08/466,662
PRIORITY FILING DATE: 1995-06-06
PRIORITY APPLICATION NUMBER: 08/256,848
PRIORITY FILING DATE: 1994-10-21
PRIORITY APPLICATION NUMBER: 09/360,934
PRIORITY FILING DATE: 1999-07-26
PRIORITY APPLICATION NUMBER: 08/471,491
PRIORITY FILING DATE: 1995-06-06
PRIORITY APPLICATION NUMBER: FCT/EP93/00472
PRIORITY FILING DATE: 1993-03-02
PRIORITY APPLICATION NUMBER: FCT/EP93/00158
PRIORITY FILING DATE: 1993-01-25
PRIORITY APPLICATION NUMBER: FI 92 A 000052
PRIORITY FILING DATE: 1992-03-02
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 5925
TYPE: DNA

```

```

; ORGANISM: Helicobacter pylori
US-09-921-157-4

Query Match      100.0%   Score 685   DB 3: Length 5925,
Best Local Similarity 100.0%   Prod. No. 3.6e-153, 7443
Matches 685; Conservative 0; Mismatches 0

```

[illegible]

RESULT 3
 US-10-335-977-3871
 ? Sequence 3871, Application US/10335977
 ? Publication No. US20040052799A1
 ?
 ? GENERAL INFORMATION:
 ?
 ? APPLICANT: DOUGLAS SMITH et al
 ?
 ? TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
 ? RELATING TO HELICOBACTER PYLORI FOR
 ? DIAGNOSTICS AND THERAPEUTICS
 ?
 ? NUMBER OF SEQUENCES: 10031
 ?
 ? CORRESPONDENCE ADDRESS:
 ?
 ? ADDRESSEE: LAHIVE & COCKFIELD
 ? STREET: 28 State Street
 ? CITY: Boston
 ? STATE: Massachusetts
 ? COUNTRY: USA
 ? ZIP: 02109-1875

Tue Mar 7 12:58:26 2006

us-09-360-685c-26.rnpbm

Page 4

```
Db 2353 TCATTAAGATGTGATCATCATCAAAAGATACGATTAAGTTGATATCTCAATCA 2412
Qy 121 GCGGTATCAGTGGCTTAAGCAAGCGGTAATTTCAAGTGGTGAAGCAAGCCGTAGCGAT 180
Db 2413 GCGGTATCAGTGGCTTAAGCAAGCGGTAATTTCAAGTGGTGAAGCAAGCCGTAGCGAT 2472
Qy 181 CTCMAAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCTCAAAAAAATGAAGTCTCAAT 240
Db 2473 CTCMAAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCTCAAAAAAATGAAGTCTCAAT 2532
Qy 241 GCTAGAAAAAATCTGAATAATTCATTCGTTAAGATGTTGAATGGAACCCTAGTC 300
Db 2533 ACTGABAAAAATTTCTGACATACATTCGTTAAGATGTTGAATGGAACCCTAGTC 2592
Qy 301 GGTAAAGGTTATCTCAAGCAAGCAAGCAAGCTCTTTCTAAAAAATTTGGACATCAAG 360
Db 2593 GGTAAAGGTTATCTCAAGCAAGCAAGCAAGCTCTTTCTAAAAAATTTGGACATCAAG 2652
Qy 361 AAAGAGTTGAATGCAAACTTGAAATTTCAATTAACAATTAATGAAGTCAAAAC 420
Db 2653 AAAGAGTTGAATGCAAACTTGAAATTTCAATTAACAATTAATGAAGTCAAAAC 2712
Qy 421 GAACCAATTTATGCTAAAGTTAATTAAGAAAGCAAGGCAAGCAAGCTTGAAAGA 480
Db 2713 AGCAC-----AGAA 2721
Qy 481 CCCATTACGCTCAAGTTGCTAAAAAGTTAATGCAAAAAATTTGACGACTCAATCAATA 540
Db 2722 CCCATTATTAATCTCAAGTTGCTAAAAAGTTAATGCAAAAAATTTGACGACTCAATCAATA 2781
Qy 541 GCAAGTGGTTGGGTGTGTGAGGCAAGCAAGGCGCTTCCCTTGAAAGAGCATGATAA 600
Db 2782 GCAAGTGGTTGGGTGTGTGAGGCAAGCAAGGCGCTTCCCTTGAAAGAGCATGATAA 2841
Qy 601 GTTGATGATCTCAGTAAGGTAGGCTTTCAAGGAATCAAGA 642
Db 2842 GTTGATGATCTCAGTAAGGTAGGCTTTCAAGGAATCAAGA 2893
```

```
RESULT 5
US-10-335-977-3873
; Sequence 3873, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHYVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-28975
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
```

```
TELEPHAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 3873:
SEQUENCE CHARACTERISTICS:
LENGTH: 3549 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HIPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...3549
SEQUENCE DESCRIPTION: SEQ ID NO: 3873:
US-10-335-977-3873
Query Match 69.7%; Score 477.2; DB 7; Length 3549;
Best Local Similarity 86.9%; Pred. No. 1,5e-103;
Matches 558; Conservative 0; Mismatches 33; Indels 51; Gaps 1;
Qy 1 AAAAAATGCAAAAAATTAAGATTTGACAGTTACGCAAGCAAAAAAGCACTTGAAAT 60
Db 2308 AAAAAATGCAAAAAATTAAGATTTGACAGTTACGCAAGCAAAAAAGCACTTGAAAT 2367
Qy 61 TCCGTTAAAGATGATCATCATCAAAAGTTAAGCAATTAAGTTGATATCTCAATCA 120
Db 2368 TCCATTAAGATGATCATCATCAAAAGTTAAGCAATTAAGTTGATATCTCAATCA 2427
Qy 121 GCGGTATCAGTGGCTTAAGCAAGCGGTAATTTCAAGTGGTGAAGCAAGCCGTAGCGAT 180
Db 2428 GCGGTATCAGTGGCTTAAGCAAGCGGTAATTTCAAGTGGTGAAGCAAGCCGTAGCGAT 2487
Qy 181 CTCMAAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCTCAAAAAAATGAAGTCTCAAT 240
Db 2488 CTCMAAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCTCAAAAAAATGAAGTCTCAAT 2547
Qy 241 GCTAGAAAAAATCTGAATAATTCATTCGTTAAGATGTTGAATGGAACCCTAGTC 300
Db 2548 GCTAGAAAAAATCTGAATAATTCATTCGTTAAGATGTTGAATGGAACCCTAGTC 2607
Qy 301 GGTAAAGGTTATCTCAAGCAAGCAAGCAAGCTCTTTCTAAAAAATTTGGACATCAAG 360
Db 2608 GGTAAAGGTTATCTCAAGCAAGCAAGCAAGCTCTTTCTAAAAAATTTGGACATCAAG 2667
Qy 361 AAAGAGTTGAATGCAAACTTGAAATTTCAATTAACAATTAATGAAGTCAAAAC 420
Db 2668 AAAGAGTTGAATGCAAACTTGAAATTTCAATTAACAATTAATGAAGTCAAAAC 2727
Qy 421 GAACCAATTTATGCTAAAGTTAATTAAGAAAGCAAGGCAAGCAAGCTTGAAAGA 480
Db 2728 AGCAC-----AGAA 2736
Qy 481 CCCATTACGCTCAAGTTGCTAAAAAGTTAATGCAAAAAATTTGACGACTCAATCAATA 540
Db 2737 CCCATTATTAATCTCAAGTTGCTAAAAAGTTAATGCAAAAAATTTGACGACTCAATCAATA 2796
Qy 541 GCAAGTGGTTGGGTGTGTGAGGCAAGCAAGGCGCTTCCCTTGAAAGAGCATGATAA 600
Db 2797 GCAAGTGGTTGGGTGTGTGAGGCAAGCAAGGCGCTTCCCTTGAAAGAGCATGATAA 2856
Qy 601 GTTGATGATCTCAGTAAGGTAGGCTTTCAAGGAATCAAGA 642
Db 2857 GTTGATGATCTCAGTAAGGTAGGCTTTCAAGGAATCAAGA 2898
```

```
RESULT 6
US-10-332-841-1/C
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Bpigenomics AG
```

;; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;; FILE REFERENCE: E01/1208/MO
;; CURRENT APPLICATION NUMBER: US/10/312,841
;; CURRENT FILING DATE: 2002-12-30
;; NUMBER OF SEQ ID NOS: 2
;; SEQ ID NO 1
;; LENGTH: 3673778
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;; FEATURE:
;; NAME/KEY: unsure
;; LOCATION: (3294164)
;; US-10-312-841-1

Query Match 7.4%; Score 50.6; DB 6; Length 3673778;
Best Local Similarity 47.4%; Pred. No. 3.3;
Matches 152; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 220 CAAAAAATGAAAGTCTCAATGCTAGAAAAAATCTGAATATATCAATCCGTTAAGAT 279
DB 2724190 CACTTAAAAAATCAATCAATTTATTAATAAAAAAATTAATTCACCCCACTTAAC 2724131
QY 280 GGTGTAATGGAACCTAGTCGTAATGSGTTATCTCAAGCAGAGCCCACTCTTCT 339
DB 2724130 TCATATATTTAAAAAACTCTACTTAAAAAATACATTCACTTAACCAAAATTAAT 2724071
QY 340 AAAAATCTTTGGACATCAAGAAAGTTGATGCAAACTTGAAATTTCAATTAACAT 399
DB 2724070 AAAAAACCATATTAATTAATAAAGAAATTCAAAAAATTAATAAATTAATAA 2724011
QY 400 AACAAATATGATGCAAAAAAGAACCACTTTATGCTAAAGTTATTAATAAAGAAAGG 459
DB 2724010 AAAAAATTAATAAATCTTAATTAATAAATTAATAAATTAATAAATTAATAA 2723951
QY 460 CAAGCACTAGCTTGAAGAACCACTTTATGCTCAAGTTGCTAAAGTTAATGAAAA 519
DB 2723950 CTAACTATTAACCTTCAAAAAATTAATAAATTAATAAATTAATAAATTAATAA 2723891
QY 520 ATTGACCACTCAATCAATA 540
DB 2723890 ATCAAAAACTCTATATATTA 2723870

RESULT 7
US-09-925-065A-20908

;; Sequence 20908, Application US/09925065A
;; Publication No. US20050228172A9
;; GENERAL INFORMATION:

;; APPLICANT: Mang, David G.
;; TITLE OF INVENTION: Identification and Mapping of Single
;; FILE REFERENCE: Nucleotide Polymorphisms in the Human Genome
;; CURRENT APPLICATION NUMBER: US/09/925,065A
;; CURRENT FILING DATE: 2001-08-08
;; PRIOR APPLICATION NUMBER: US 60/243,096
;; PRIOR FILING DATE: 2000-10-24
;; PRIOR APPLICATION NUMBER: US 60/252,147
;; PRIOR FILING DATE: 2000-11-20
;; PRIOR APPLICATION NUMBER: US 60/250,092
;; PRIOR FILING DATE: 2000-11-30
;; PRIOR APPLICATION NUMBER: US 60/261,766
;; PRIOR FILING DATE: 2001-01-16
;; PRIOR APPLICATION NUMBER: US 60/289,846
;; PRIOR FILING DATE: 2001-05-09
;; NUMBER OF SEQ ID NOS: 957086
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 20908
;; LENGTH: 571
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; US-09-925-065A-20908

Query Match 7.0%; Score 48.2; DB 4; Length 571;
Best Local Similarity 50.7%; Pred. No. 0.27;
Matches 116; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 221 AAAAAAATGAAAGTCTCAATGCTAGAAAAAATCTGAATATATCAATCCGTTAAGAT 280
DB 24 MAGAAAACTACGATTAACCCGATGAAAAAGGGCAAAAGTCTTCAGCACTTACAAATG 83
QY 281 GTGTGAATGGAACCTAGTCGTAATGSGTTATCTCAAGCAGAGCCCACTCTTCTA 340
DB 84 GTATGTTTGAATGAATGATTTGATTCATTAATTTGATTCACATGGCCATTAATATATG 143
QY 341 AAAAATCTTTGGACATCAAGAAAGTTGATGCAAACTTGAAATTTCAATTAACATA 400
DB 144 AAAAGATATGATGATTCATCAAGCAGCAATTAAGAAAAATTAAGATTAATAAATGCAATGAGCTAA 203
QY 401 ACATATATGACTCAAAAAAGAACCATTTATGCTTAAGTTAATAAATA 449
DB 204 AACGATATGAAAAAAGAAATTAATCAAAATGTTAAGAAAAAGATGTGAAAAA 252

RESULT 8
US-10-311-455-1916/c

;; Sequence 1916, Application US/10311455
;; Publication No. US20030143606A1
;; GENERAL INFORMATION:

;; APPLICANT: OLEK, Alexander
;; APPLICANT: PIEPENBROCK, Christian
;; APPLICANT: BERLIN, Kurt
;; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by De-
;; FILE REFERENCE: 5013.1014
;; CURRENT APPLICATION NUMBER: US/10/311,455
;; CURRENT FILING DATE: 2002-12-16
;; PRIOR APPLICATION NUMBER: PCT/EP01/07537
;; PRIOR FILING DATE: 2001-07-02
;; PRIOR APPLICATION NUMBER: DE 10032529.7
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: DE 10043826.1
;; PRIOR FILING DATE: 2000-09-01
;; NUMBER OF SEQ ID NOS: 2424
;; SEQ ID NO 1916
;; LENGTH: 12138
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;; US-10-311-455-1916

Query Match 7.0%; Score 48; DB 6; Length 12138;
Best Local Similarity 47.5%; Pred. No. 1.1;
Matches 174; Conservative 0; Mismatches 190; Indels 2; Gaps 1;

QY 161 TAGCAAGCGTTAGCCGATCTCAAAAAATTTCTCAAGAGCAATGCGCCCAAGAGCTC 220
DB 9844 TATATAAATCACTAATATTAATAAATCTCTTAATTAATAAATTAATCAAAACAACT 9785
QY 221 AAAAAAATGAAAGTCTCAATGCTAAGAAAAATCTGAATATATCAATCCGTTAAGAT 280
DB 9784 TCAAAAAATTCATCTAATCAATTAATAAATAAATCACTTATTAATTAACATTAACAAAC 9725
QY 281 GTGTGAATGGAACCTAGTCGTAATGSGTTATCTCAAGCAGAGCCCACTCTTCTA 340
DB 9724 CCCCTCTTAATTAATAATTTCTCAATCAATTTATTTTAAAAATTAACAAACCACTTCA 9665
QY 341 AAAAATCTTTGGACATCAAGAAAGTTGATGCAAACTTGAAATTTCAATTAACATA 400
DB 9664 AAAAATATTTTCAATTTCAAAAAATCAAAAAATTAATAAATTAATTAATATATATA 9605
QY 401 ACAAT--AATGACCTCAAAAAAGAACCATTTATGCTTAATTAATAAAGAAAGCAGG 458
DB 9604 AAAATCAACAAACAAACAAATTAACAAAAAATTAATAAATAAATAAATAAATAATCTAA 9545

QY 459 GCAAGCAGCTAGCCTTGAAGAACCCATTAGGCTCAAGTTGCTAAAAAGTAAATGCANA 518
DB 9544 AAAACCACTATCAAAAAAABAAAAAACAATTAATATACCAAAAAAATCTTAA 9485
QY 519 AATTGA 524
DB 9484 TTTTAA 9479

RESULT 9
US-10-240-453-210/c
Sequence 210, Application US/10240453
Publication No. US20030146326A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
TITLE OF INVENTION: Transcription
TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
FILE REFERENCE: 5013.1009
CURRENT FILING DATE: 2002-10-02
PRIOR APPLICATION NUMBER: US/10/240,453
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: PCT/EP01/03973
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 350
SEQ ID NO 210
LENGTH: 12138
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-210

Query Match 7.0%; Score 48; DB 6; Length 12138;
Best Local Similarity 47.5%; Pred. No. 1.1; Indels 2; Gaps 1;
Matches 174; Conservative 0; Mismatches 190; Indels 2; Gaps 1;
QY 161 TAGAGCAGCGTTACCGATCTCAAAAATTTCTCAAGAGCAATTGGCCCAACAGCTC 220
DB 9844 TATTAATATCACTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9785
QY 221 AAAAAATGAAGCTCAATGCTGAGAAAAAATGAAATATATCAATCCGTTAAGATG 280
DB 9784 TCAAAAAATTCATCTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9725
QY 281 GTGTGAATGAAACCTGTGCGTAAATGCTATCTCAAGCAGACCAACTCTTCA 340
DB 9724 CCCCTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9665
QY 341 AAAATTTTCGACATCAAGAAAGATTGCAAACTTGAAATTTCAATTAACATA 400
DB 9664 AAAATTTTCATTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9605
QY 401 ACAT--AATGACTCAAAAGCAACCACTTTTGTAAAGTAAATTAATTAATTAATTAATTA 9545
DB 9604 AAAATTAACAAACCAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9545
QY 459 GCAAGCAGCTAGCCTTGAAGAACCCATTAGGCTCAAGTTGCTAAAAAGTAAATGCANA 518
DB 9544 AAAACCACTATCAAAAAAABAAAAAACAATTAATATACCAAAAAAATCTTAA 9485
QY 519 AATTGA 524

DB 9484 TTTTAA 9479

RESULT 10
US-10-312-841-2/c
Sequence 2, Application US/10312841
Publication No. US20030186277A1
GENERAL INFORMATION:
APPLICANT: Epigenomics AG
TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MR
FILE REFERENCE: E01/1208/MO
CURRENT APPLICATION NUMBER: US/10/312,841
CURRENT FILING DATE: 2002-12-30
NUMBER OF SEQ ID NOS: 2
SEQ ID NO 2
LENGTH: 3673778
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
NAME/KEY: unsure
LOCATION: (379615)
US-10-312-841-2

Query Match 7.0%; Score 47.8; DB 6; Length 3673778;
Best Local Similarity 46.5%; Pred. No. 1.5; Indels 0; Gaps 0;
Matches 154; Conservative 0; Mismatches 177; Indels 0; Gaps 0;
QY 183 CAAAAATTTCTCAAGAGCAATTGGCCCAACCAAGCTCAAAAAATGAAGTCTCAATGC 242
DB 3055754 CAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3055695
QY 243 TAGAAAAATCTGAATATATATCAATCCGTTAAGATGCTGAAATGAAACCCCTAGTCGG 302
DB 3055694 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3055635
QY 303 TAATGGTATATCTCAAGAGCAAGCCCAACTCTTCTAAAACTTTTGGACATCAAGA 362
DB 3055634 ATTAACAATCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3055575
QY 363 AGATTAATGCAAACTGGAATTTCAATTAACATTAATTAATTAATTAATTAATTAATTA 422
DB 3055574 CAATCTAATTAACAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3055515
QY 423 ACCATTTTATGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 482
DB 3055514 AAAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3055455
QY 483 CATTAGGCTCAAGTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 513
DB 3055454 AATTTAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3055424

RESULT 11
US-10-311-455-1931/c
Sequence 1931, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1

PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1931
LENGTH: 14006
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
NAME/KEY: unsure
LOCATION: 8289, 8310, 8313
OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1931

Query Match
Best Local Similarity 44.7%; Score 47; DB 6; Length 14006;
Pred. No. 2.1; Mismatches 275; Indels 1; Gaps 1;
Matches 223; Conservative 0;

42 AAAAAAGCAGCTTGAATAATTCGTTAAAGATGATCATCAATCAAAAGGTAAAGCGATTA 101
6018 AAAAAAAAAAATTAACCACTTAACCAAAATCAAACTAAATTAATTAATTAATTAATTA 5959
102 AGTTATATCTCAATCAAGCGGTATCAAGTGGCTAAGCAACGGGTATTCAGTAGGGT 161
5958 ATTATATTAATAAATCAAA-CACTAACAAATAAATTAATTAATTAATTAATTAATTA 5900
162 AAGAGCAAGGTAGCGGATCTCAAAATTTCTCAAGAGCAATTTGCCCAACAGCTCA 221
5899 ATTATCTGACTTCATCAATCAAAATTTTAATAATTAATTAATTAATTAATTAATTA 5840
222 AAAAAATGAAGCTCAATGCTGAATAAATCTGAATATATCAATCGTTAAGATGG 281
5839 ATTATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5780
282 TGTGAATGACACCTTAAGTGGTATCTCAAGCAAGCAACTCTTTCTTA 341
5779 CCAAAATACATATTAACCAAAATCTTAATAATTAATTAATTAATTAATTAATTA 5720
342 AACTTTTGGACATCAAGAAAGTTGAATGCAAACTTGAATTTCAATTAACATAA 401
5719 CAATTTCTAATAAATTTCTAATAACACACACAAATTAATTAATTAATTAATTA 5660
402 CAATATGAGCTCAAAAGCAACCCATTATGCTAAGTTAATAAATAAAGACGGGCA 461
5659 CAACACTATTTCTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5600
462 AGCAGTACCTTGAAGAACCATTTACCTCAAGTTGCTAATAAAGTAAATGCAAAAT 521
5599 AATTAACCTTTCTAATCAACATTAATAAATAAATAAATAAATAAATAAATAAATAA 5540
522 TGACGACTCAATCAATA 540
5539 AAAAAATTCACAAATCTAATA 5521

RESULT 12
US-10-311-455-1993/c
Sequence 1993, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
PRIOR FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1

PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1993
LENGTH: 5542
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1993

Query Match
Best Local Similarity 48.2%; Score 46.8; DB 6; Length 5542;
Pred. No. 1.6; Mismatches 142; Indels 0; Gaps 0;
Matches 132; Conservative 0;

180 TCTCAAAATTTCTCAAGAGCAATTTGCCCAACAGCTCAAAAAATGAAGTCTCA 239
4485 TCACATATATTCAAAAATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4426
240 TCTTGAATAAATTCGAAATTAATTAATTCGTTAAGATGAGTGAATGCAACCTTGT 299
4425 TCATTAACCTTAATTAATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4366
300 CGGTATGAGTATCTCAAGCAAGCAACCACTTTCTTAATAAATTTTGGACATCA 359
4365 CAAAAAATAAATAAATCAACCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4306
360 GAAAGTTGAATGCAAACTTGAATTTCAATTAATTAATTAATTAATTAATTAATTAATTA 419
4305 AACGTTTATACATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4246
420 CGAACCCATTATGCTAAGTTAATTAATAAAGAA 453
4245 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4212

RESULT 13
US-10-311-455-71/c
Sequence 71, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
PRIOR FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 71
LENGTH: 9964
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-71

Query Match
Best Local Similarity 45.4%; Score 46.8; DB 6; Length 9964;
Pred. No. 2; Mismatches 292; Indels 5; Gaps 2;
Matches 247; Conservative 0;

1 AAAAAAGCAGCAATTAAGATTTTCAGCAAGTAAAGCAAGCAAAAGCACTTGAAT 60
9281 AAAAAAAGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9222
61 TCCGTTAAGATGATCATCAATCAAAAGGTAAAGGATTAAGTTATTAATCAATCA 120

Db 9221 AAAAAAAAAAATCAAAAAAAAAAAAAAAAAAATTAACACAAACATCAATATA 9162
Qy 121 GCGGTATGAGTGGCTAAAGCAAGGGTATTTCTAGTAGGTAGCAAGCTTAGCCGAT 180
Db 9161 TATTATTAATTTTAAATTA-----ATTAAATTAATTAACAAATCTTAATTAATTA 9106
Qy 181 CTCAAAAATTTCTCAAGAGCAATGGCCCAACAGCTCAAAAAATGAAGTCTCAT 240
Db 9105 AAAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9046
Qy 241 GCTAGAAAAATCTGAATATTAATCAATCGCTTAAGATGTGTGAATGAAACCTGATC 300
Db 9045 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8986
Qy 301 GGTATGGTTATCTCAAGCAAGCCCAACTCTTTCTAAAACTTTTGGACATCAAG 360
Db 8985 AAAAAAAAAAATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8926
Qy 361 AAAGAGTTGAATGCAAACTTGAATTTCAATTAATTAATTAATTAATTAATTA 420
Db 8925 ATCGAAAAATTA-----ATAATTAATTAATTAATTAATTAATTAATTAATTA 8867
Qy 421 GAACCATTTATGCTAAAGTTAATTAATTAATTAATTAATTAATTAATTAATTA 8807
Db 8866 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8807
Qy 481 CCCATTTAGCTCAAGTGTCTAAAGTTAATTAATTAATTAATTAATTAATTA 540
Db 8806 TAAAAATTTATCAAAATTTATCAATTAATTAATTAATTAATTAATTAATTA 8747
Qy 541 GCAA 544
Db 8746 TAAA 8743

RESULT 14
US-10-425-115-38710/c
/ Sequence 38710, Application US/10425115
/ Publication No. US20040214272A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovacic, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(53222)B
/ CURRENT APPLICATION NUMBER: US/10/425,115
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 369326
/ SEQ ID NO 38710
/ LENGTH: 921
/ TYPE: DNA
/ ORGANISM: Zea mays
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(921)
/ OTHER INFORMATION: unsure at all n locations
/ FEATURE:
/ OTHER INFORMATION: Clone ID: MRT4577_135307C.1
US-10-425-115-38710

Query Match 6.8%; Score 46.4; DB 8; Length 921;
Best Local Similarity 42.8%; Pred. No. 0.9;
Matches 223; Conservative 0; Mismatches 311; Indels 0; Gaps 0;
Qy 1 AAAAAAGCAAAATTAAGATTTCAAGCAAGTAAAGCAAAAGCAAGCTTGAAT 60
Db 851 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 792
Qy 61 TCCGTTAAGATGTATCATCAATCAAAAGTAAAGCAATTAATTAATTAATTAATCA 120
Db 791 AATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 732

Qy 121 GCGGTATGAGTGGCTAAAGCAAGGGTATTTCTAGTAGGTAGCAAGCTTAGCCGAT 180
Db 731 AAAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 672
Qy 181 CTCAAAAATTTCTCAAGAGCAATGGCCCAACAGCTCAAAAAATGAAGTCTCAT 240
Db 671 AAAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 612
Qy 241 GCTAGAAAAATCTGAATATTAATCAATCGCTTAAGATGTGTGAATGAAACCTGATC 300
Db 611 AAAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 552
Qy 301 GGTATGGTTATCTCAAGCAAGCCCAACTCTTTCTAAAACTTTTGGACATCAAG 360
Db 551 AAAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 492
Qy 361 AAAGAGTTGAATGCAAACTTGAATTTCAATTAATTAATTAATTAATTAATTA 420
Db 491 AAAAAATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 432
Qy 421 GAACCATTTATGCTAAAGTTAATTAATTAATTAATTAATTAATTAATTAATTA 480
Db 431 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 372
Qy 481 CCCATTTAGCTCAAGTGTCTAAAGTTAATTAATTAATTAATTAATTAATTA 540
Db 371 TAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 312
Qy 541 GCAA 544
Db 311 AAAA 308

RESULT 15
US-10-239-676-132/c
/ Sequence 132, Application US/10239676
/ Publication No. US20030082609A1
/ GENERAL INFORMATION:
/ APPLICANT: Olex, Alexander
/ APPLICANT: Prie, Alexander
/ APPLICANT: Berlin, Kurt
/ TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
/ FILE REFERENCE: 5013.1003
/ CURRENT APPLICATION NUMBER: US/10/239,676
/ CURRENT FILING DATE: 2002-09-24
/ PRIOR APPLICATION NUMBER: PCT/EP01/03968
/ DE 10019058.8
/ DE 10032529.7
/ DE 10043826.1
/ PRIOR FILING DATE: 2001-04-06
/ 2000-04-06
/ 2000-04-07
/ 2000-06-30
/ 2000-09-01
/ NUMBER OF SEQ ID NOS: 228
/ SEQ ID NO 132
/ LENGTH: 5127
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
/ NAME/KEY: unsure
/ LOCATION: (1976, 1981..1982, 1984, 1989..1990, 1995..1996, 5032, 5059)
US-10-239-676-132

Query Match 6.8%; Score 46.4; DB 5; Length 5127;
Best Local Similarity 47.9%; Pred. No. 1.9;
Matches 163; Conservative 0; Mismatches 176; Indels 1; Gaps 1;
Qy 184 AAAAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCAAAAAATTAATTAATTAATCA 243


```
Db      1173 AACAAATTACTAAAAAAATCTCGAAAAAAATAAACTCAAAAAAAATTAATTTT 1114
Qy      244 AGAAAAAATCTGAATATATCATCCGTTAAGATGCTGATGAACCTAGTCGT 303
Db      1113 AAAAAACGATCTCTAAAAAAATTTTATATAATTACGAAACGAAATATTCTAANA 1054
Qy      304 AATGGTTATCTCAAGCAGAGCCACACTCTTTCTAAAAAATTTTCGGACATCAGAAA 363
Db      1053 AAAAAAATTTACGAAAAAAATTTTACCAGAAAAAAATTTCTAAAAAAATTTT 994
Qy      364 GAGTTGATCGAAACTTGGAATTTCAATTAACAATTAATGAGTCAAAAAAG-A 422
Db      993 TCCAAAAATACTTAATCTTAACAAAAAAATTTCTAAAAAAATACTAAAAAAACGAA 934
Qy      423 ACCCAATTATGCTAAAGTTAATTAAGAAAGCAGGCAAGCACTAGCTTGAAGACC 482
Db      933 AACCGAATAAAAAAATTCAAAAAAATTTCTAAATTTAACACTCTTAAAAAAA 874
Qy      483 CATTTACGCTCAAGTTGCTAAAAAAGGTAATGCAAAATT 522
Db      873 CATTTAATTTAACCATCTAAAAAAATTTAAAAAACATT 834
```

Search completed: March 6, 2006, 22:25:26
Job time : 786 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2006, 21:38:13 ; Search time 183 Seconds
(without alignments)
6653.714 Million cell updates/sec

Title: US-09-360-685C-26

Perfect score: 685
Sequence: 1 aaaaatgcgaataaagga.....caatcaagcgatcagaag 685

Scoring table: IDENTITY_NIC
Gapop 10.0 , Gapept 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1 COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5 COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6H COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PTUS COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PT COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	685	100.0	5925	3	US-08-470-260-4 Sequence 4, Appl1
2	685	100.0	5925	3	US-08-471-491-4 Sequence 4, Appl1
3	685	100.0	5925	3	US-08-466-662-4 Sequence 4, Appl1
4	685	100.0	10299	2	US-08-477-451-1 Sequence 1, Appl1
5	685	100.0	10299	2	US-08-477-451-1 Sequence 5, Appl1
6	685	100.0	19932	2	US-08-477-451-25 Sequence 25, Appl1
7	566.6	82.7	815	3	US-08-893-010-3 Sequence 3, Appl1
8	485.2	70.8	4821	2	US-08-053-614-3 Sequence 3, Appl1
9	485.2	70.8	4821	2	US-08-316-397B-3 Sequence 3, Appl1
10	485.2	70.8	4821	2	US-09-034-306-3 Sequence 3, Appl1
11	485.2	70.8	4821	6	PCT-US93-09782-3 Sequence 3, Appl1
12	485.2	70.8	4821	6	PCT-US93-09782-3 Sequence 3, Appl1
13	277.2	40.5	3648	2	US-08-053-614-1 Sequence 1, Appl1
14	277.2	40.5	3648	2	US-08-316-397B-1 Sequence 1, Appl1
15	277.2	40.5	3648	2	US-09-034-306-1 Sequence 1, Appl1
16	277.2	40.5	3648	2	US-09-259-437-1 Sequence 1, Appl1
17	277.2	40.5	3648	6	PCT-US93-09782-1 Sequence 1, Appl1
18	56	8.2	7218	2	US-08-232-463-14 Sequence 14, Appl1
19	45.6	6.7	849	3	US-09-248-796A-2496 Sequence 7183, Ap
20	44.8	6.5	1128	3	US-09-569-348-5 Sequence 5, Appl1
21	44.8	6.5	1131	9	US-08-795-475-5 Patent No. 5180810
22	44.8	6.5	1332	3	US-08-325-278B-5 Sequence 5, Appl1
23	44.8	6.5	1332	3	US-08-325-278B-5 Sequence 5, Appl1
24	44.8	6.5	1784	9	US-08-325-278B-5 Patent No. 5180810

25	44.4	6.5	612	3	US-09-902-540-1357 Sequence 1357, Ap
26	44	6.4	750	3	US-09-248-796A-499 Sequence 499, App
27	43.8	6.4	642	3	US-09-248-796A-4755 Sequence 4755, Ap
28	43.8	6.4	1039	3	US-09-902-540-1280 Sequence 1280, Ap
29	43.2	6.3	765	3	US-08-914-479A-3 Sequence 3, Appl1
30	43.2	6.3	918	3	US-08-937-271-9 Sequence 9, Appl1
31	42	6.1	606	3	US-09-248-796A-2496 Sequence 2496, Ap
32	41.6	6.1	855	3	US-08-914-479A-5 Sequence 5, Appl1
33	41.6	6.1	205183	3	US-09-949-016-17009 Sequence 17009, A
34	41.4	6.0	152070	3	US-09-949-016-135879 Sequence 135879, A
35	41.2	6.0	601	3	US-09-949-016-135879 Sequence 135879, A
36	41.2	6.0	601	3	US-09-949-016-135996 Sequence 135996, A
37	41.2	6.0	601	3	US-09-949-016-136113 Sequence 136113, A
38	41.2	6.0	193555	3	US-09-949-016-15553 Sequence 15553, A
39	41.2	6.0	193555	3	US-09-949-016-15554 Sequence 15554, A
40	41.2	6.0	193555	3	US-09-949-016-15554 Sequence 15555, A
41	40.6	5.9	6744	2	US-08-119-125A-2 Sequence 2, Appl1
42	40.4	5.9	30549	3	US-09-134-001C-322 Sequence 322, App
43	40.2	5.9	212	3	US-09-270-767-30271 Sequence 30271, A
44	40.2	5.9	1055	3	US-09-806-708B-23 Sequence 23, Appl1
45	40.2	5.9	1967	3	US-09-270-767-14158 Sequence 14158, A

ALIGNMENTS

RESULT 1
US-08-470-260-4
; Sequence 4, Application US/08470260
; Patent No. 6077706
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappoli, Rino
TITLE OF INVENTION: Helicobacter Pylori Proteins Useful
TITLE OF INVENTION: for Vaccines and Diagnostics
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,260
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,848
FILING DATE: 21-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0316.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 5925 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-470-260-4

Query Match 100.0%; Score 685; DB 3; Length 5925;
Best Local Similarity 100.0%; Pred. No. 2,4e-158;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AAAATGGCAAAATTAAGATTTCAGCAAGGTACGCAAGCAAAAGGACCTTGAAT 60
DB 2782 AAAAATGGCAAAATTAAGATTTCAGCAAGGTACGCAAGCAAAAGGACCTTGAAT 2841
QY 61 TCCGTTAAAGATGTGATCATCAATCAAAAGGTAAAGGTTAAAGTTGATCTCAATCA 120
DB 2842 TCCGTTAAAGATGTGATCATCAATCAAAAGGTAAAGGTTAAAGTTGATCTCAATCA 2901
QY 121 GCGGTATCAGTGGCTAAAGCAAGGGTGAATTCAGTAGGTAGAGCAAGGTTAGCCAT 180
DB 2902 GCGGTATCAGTGGCTAAAGCAAGGGTGAATTCAGTAGGTAGAGCAAGGTTAGCCAT 2961
QY 181 CTCAAAAATTTCTCAAGAGGCAATGGCCCAACAGCTCAAAAAATGAAGTCTCAAT 240
DB 2962 CTCAAAAATTTCTCAAGAGGCAATGGCCCAACAGCTCAAAAAATGAAGTCTCAAT 3021
QY 241 GCTAGAAAAAATCTGAATATATCAATCCGTTAAAGATGGTGAATGGAACCTTACAT 300
DB 3022 GCTAGAAAAAATCTGAATATATCAATCCGTTAAAGATGGTGAATGGAACCTTACAT 3081
QY 301 GGTATGGGTTATCTCAAGAGGCAAGCCCAACTCTTTCTAAAACTTTTGGACATCAAG 360
DB 3082 GGTATGGGTTATCTCAAGAGGCAAGCCCAACTCTTTCTAAAACTTTTGGACATCAAG 3141
QY 361 AAAGAGTTGAATGCAAAATCTGGAATTTCAATACAAATTAATCAATATGACTCAAAAAC 420
DB 3142 AAAGAGTTGAATGCAAAATCTGGAATTTCAATACAAATTAATCAATATGACTCAAAAAC 3201
QY 421 GAACCCATTATCTGAAGTTATATTAAGAAAGAGGCAAGGCAAGGCTTGAAGAA 480
DB 3202 GAACCCATTATCTGAAGTTATATTAAGAAAGAGGCAAGGCAAGGCTTGAAGAA 3261
QY 481 CCCATTACGCTCAAGTTGCTTAAAGGTAAATGCAAAAAATTTGACGACTCATCAATA 540
DB 3262 CCCATTACGCTCAAGTTGCTTAAAGGTAAATGCAAAAAATTTGACGACTCATCAATA 3321
QY 541 GCAAGTGGTTGGGTGTGTGAGGCAAGCGGGCTTCCCTTTGAAAAAGCATGATPAA 600
DB 3322 GCAAGTGGTTGGGTGTGTGAGGCAAGCGGGCTTCCCTTTGAAAAAGCATGATPAA 3381
QY 601 GTTATATCTCAAGTGAAGTGGCTTTCAAGGAATCAAGAAATTTGCTCAAGAAATTTGAC 660
DB 3382 GTTATATCTCAAGTGAAGTGGCTTTCAAGGAATCAAGAAATTTGCTCAAGAAATTTGAC 3441
QY 661 AATCTCAATCAAGCGGTATCAAGAG 685
DB 3442 AATCTCAATCAAGCGGTATCAAGAG 3466
```

RESULT 2
US-08-471-491-4
Sequence 4, Application US/08471491B

GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
FILE REFERENCE: CHIR0044
CURRENT APPLICATION NUMBER: US/08/471,491B
CURRENT FILING DATE: 1995-06-06
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 5925

TYPE: DNA
ORGANISM: Helicobacter pylori
US-08-471-491-4
Query Match 100.0%; Score 685; DB 3; Length 5925;
Best Local Similarity 100.0%; Pred. No. 2,4e-158;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AAAATGGCAAAATTAAGATTTCAGCAAGGTACGCAAGCAAAAGGACCTTGAAT 60
DB 2782 AAAAATGGCAAAATTAAGATTTCAGCAAGGTACGCAAGCAAAAGGACCTTGAAT 2841
QY 61 TCCGTTAAAGATGTGATCATCAATCAAAAGGTAAAGGTTAAAGTTGATCTCAATCA 120
DB 2842 TCCGTTAAAGATGTGATCATCAATCAAAAGGTAAAGGTTAAAGTTGATCTCAATCA 2901
QY 121 GCGGTATCAGTGGCTAAAGCAAGGGTGAATTCAGTAGGTAGAGCAAGGTTAGCCAT 180
DB 2902 GCGGTATCAGTGGCTAAAGCAAGGGTGAATTCAGTAGGTAGAGCAAGGTTAGCCAT 2961
QY 181 CTCAAAAATTTCTCAAGAGGCAATGGCCCAACAGCTCAAAAAATGAAGTCTCAAT 240
DB 2962 CTCAAAAATTTCTCAAGAGGCAATGGCCCAACAGCTCAAAAAATGAAGTCTCAAT 3021
QY 241 GCTAGAAAAAATCTGAATATATCAATCCGTTAAAGATGGTGAATGGAACCTTACAT 300
DB 3022 GCTAGAAAAAATCTGAATATATCAATCCGTTAAAGATGGTGAATGGAACCTTACAT 3081
QY 301 GGTATGGGTTATCTCAAGAGGCAAGCCCAACTCTTTCTAAAACTTTTGGACATCAAG 360
DB 3082 GGTATGGGTTATCTCAAGAGGCAAGCCCAACTCTTTCTAAAACTTTTGGACATCAAG 3141
QY 361 AAAGAGTTGAATGCAAAATCTGGAATTTCAATACAAATTAATCAATATGACTCAAAAAC 420
DB 3142 AAAGAGTTGAATGCAAAATCTGGAATTTCAATACAAATTAATCAATATGACTCAAAAAC 3201
QY 421 GAACCCATTATCTGAAGTTATATTAAGAAAGAGGCAAGGCAAGGCTTGAAGAA 480
DB 3202 GAACCCATTATCTGAAGTTATATTAAGAAAGAGGCAAGGCAAGGCTTGAAGAA 3261
QY 481 CCCATTACGCTCAAGTTGCTTAAAGGTAAATGCAAAAAATTTGACGACTCATCAATA 540
DB 3262 CCCATTACGCTCAAGTTGCTTAAAGGTAAATGCAAAAAATTTGACGACTCATCAATA 3321
QY 541 GCAAGTGGTTGGGTGTGTGAGGCAAGCGGGCTTCCCTTTGAAAAAGCATGATPAA 600
DB 3322 GCAAGTGGTTGGGTGTGTGAGGCAAGCGGGCTTCCCTTTGAAAAAGCATGATPAA 3381
QY 601 GTTATATCTCAAGTGAAGTGGCTTTCAAGGAATCAAGAAATTTGCTCAAGAAATTTGAC 660
DB 3382 GTTATATCTCAAGTGAAGTGGCTTTCAAGGAATCAAGAAATTTGCTCAAGAAATTTGAC 3441
QY 661 AATCTCAATCAAGCGGTATCAAGAG 685
DB 3442 AATCTCAATCAAGCGGTATCAAGAG 3466
```

RESULT 3
US-08-466-662-4
Sequence 4, Application US/08466662B

GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
FILE REFERENCE: CHIR0057
CURRENT APPLICATION NUMBER: US/08/466,662B
CURRENT FILING DATE: 1995-06-06
NUMBER OF SEQ ID NOS: 8

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 5925
TYPE: DNA
ORGANISM: Helicobacter pylori
US-08-466-662-4

Query Match 100.0%; Score 685; DB 3; Length 5925;
Best Local Similarity 100.0%; Pred. No. 2.4e-158; Indels 0; Gaps 0;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AAAAATGGCAAAAATAGAGATTTCAGCAAGTAAACCAAGCAAAAAGCCACTTGAAT 60
DB 2782 AAAAATGGCAAAAATAGAGATTTCAGCAAGTAAACCAAGCAAAAAGCCACTTGAAT 2841
QY 61 TCCGTTAAAGTGTATCATCATCAAAAAGTAAAGGTAAGTGTATCATCAATCA 120
DB 2842 TCCGTTAAAGTGTATCATCATCAAAAAGTAAAGGTAAGTGTATCATCAATCA 2901
QY 121 GCGGTATCAGTGGCTAAGCAAGGGTGAATTCAGTAGGGTGAAGCAAGCTTAAAGCCAT 180
DB 2902 GCGGTATCAGTGGCTAAGCAAGGGTGAATTCAGTAGGGTGAAGCAAGCTTAAAGCCAT 2961
QY 181 CTCAAAAATTTCTCAAAAGGCAATTTGGCCCAACAAGCTCAAAAAATGAAGTCTCAAT 240
DB 2962 CTCAAAAATTTCTCAAAAGGCAATTTGGCCCAACAAGCTCAAAAAATGAAGTCTCAAT 3021
QY 241 GCTAGAAAAAATCTGAATATATTCATTCGTTAAGATGTGTGAATGAAACCTTAGTC 300
DB 3022 GCTAGAAAAAATCTGAATATATTCATTCGTTAAGATGTGTGAATGAAACCTTAGTC 3081
QY 301 GGTAAATGGTTATCTCAAGCAAGGCAAGCTCTTTTAAACCTTTTGGACATCAAG 360
DB 3082 GGTAAATGGTTATCTCAAGCAAGGCAAGCTCTTTTAAACCTTTTGGACATCAAG 3141
QY 361 AAAGAGTTGAATGCAAAAATCTGAAATTTCAATTAACAATATATGATGCTCAAAAAC 420
DB 3142 AAAGAGTTGAATGCAAAAATCTGAAATTTCAATTAACAATATATGATGCTCAAAAAC 3201
QY 421 GAACCATTTTATGCTAAATTAATTAAGAAAGCAAGGCAAGCAAGCTTGAAGAA 480
DB 3202 GAACCATTTTATGCTAAATTAATTAAGAAAGCAAGGCAAGCAAGCTTGAAGAA 3261
QY 481 CCCATTATAGCTCAAGTGTCTAAAGGTAATGCAAAAATGACCCGACTCAATCAATA 540
DB 3262 CCCATTATAGCTCAAGTGTCTAAAGGTAATGCAAAAATGACCCGACTCAATCAATA 3321
QY 541 GCAAGTGTGTTGGTGTGTGTAAGGCAAGGCGGCTTCCCTTTGAAAAGGCAATGAA 600
DB 3322 GCAAGTGTGTTGGTGTGTGTAAGGCAAGGCGGCTTCCCTTTGAAAAGGCAATGAA 3381
QY 601 GTTATGATCTCAGTAAGGTAGGGCTTTCAAGGAAATCAAGAAATGAGCTCAAGAAATGAC 660
DB 3382 GTTATGATCTCAGTAAGGTAGGGCTTTCAAGGAAATCAAGAAATGAGCTCAAGAAATGAC 3441
QY 661 AATTCATCAAGCCGATCAAG 685
DB 3442 AATTCATCAAGCCGATCAAG 3466
```

RESULT 4
US-08-477-451-1
Sequence 1, Application US/08477451
Patent No. 5928865
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter Pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSES: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA

COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335,002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10299 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-477-451-1

Query Match 100.0%; Score 685; DB 2; Length 10299;
Best Local Similarity 100.0%; Pred. No. 2.8e-158; Indels 0; Gaps 0;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AAAAATGGCAAAAATAGAGATTTCAGCAAGTAAACCAAGCAAAAAGCCACTTGAAT 60
DB 7156 AAAAATGGCAAAAATAGAGATTTCAGCAAGTAAACCAAGCAAAAAGCCACTTGAAT 7215
QY 61 TCCGTTAAAGTGTATCATCATCAAAAAGTAAAGGTAAGTGTATCATCAATCA 120
DB 7216 TCCGTTAAAGTGTATCATCATCAAAAAGTAAAGGTAAGTGTATCATCAATCA 7275
QY 121 GCGGTATCAGTGGCTAAGCAAGGGTGAATTCAGTAGGGTGAAGCAAGCTTAAAGCCAT 180
DB 7276 GCGGTATCAGTGGCTAAGCAAGGGTGAATTCAGTAGGGTGAAGCAAGCTTAAAGCCAT 7335
QY 181 CTCAAAAATTTCTCAAAAGGCAATTTGGCCCAACAAGCTCAAAAAATGAAGTCTCAAT 240
DB 7336 CTCAAAAATTTCTCAAAAGGCAATTTGGCCCAACAAGCTCAAAAAATGAAGTCTCAAT 7395
QY 241 GCTAGAAAAAATCTGAATATATTCATTCGTTAAGATGTGTGAATGAAACCTTAGTC 300
DB 7396 GCTAGAAAAAATCTGAATATATTCATTCGTTAAGATGTGTGAATGAAACCTTAGTC 7455
QY 301 GGTAAATGGTTATCTCAAGCAAGGCAAGCTCTTTTAAACCTTTTGGACATCAAG 360
DB 7456 GGTAAATGGTTATCTCAAGCAAGGCAAGCTCTTTTAAACCTTTTGGACATCAAG 7515
QY 361 AAAGAGTTGAATGCAAAAATCTGAAATTTCAATTAACAATATATGATGCTCAAAAAC 420
DB 7516 AAAGAGTTGAATGCAAAAATCTGAAATTTCAATTAACAATATATGATGCTCAAAAAC 7575
QY 421 GAACCATTTTATGCTAAATTAATTAAGAAAGCAAGGCAAGCAAGCTTGAAGAA 480
DB 7576 GAACCATTTTATGCTAAATTAATTAAGAAAGCAAGGCAAGCAAGCTTGAAGAA 7635
QY 481 CCCATTATAGCTCAAGTGTCTAAAGGTAATGCAAAAATGACCCGACTCAATCAATA 540
DB 7636 CCCATTATAGCTCAAGTGTCTAAAGGTAATGCAAAAATGACCCGACTCAATCAATA 7695
QY 541 GCAAGTGTGTTGGTGTGTGTAAGGCAAGGCGGCTTCCCTTTGAAAAGGCAATGAA 600
DB 7696 GCAAGTGTGTTGGTGTGTGTAAGGCAAGGCGGCTTCCCTTTGAAAAGGCAATGAA 7755
QY 601 GTTATGATCTCAGTAAGGTAGGGCTTTCAAGGAAATCAAGAAATGAGCTCAAGAAATGAC 660
```

Tue Mar 7 12:58:26 2006

us-09-360-685c-26.rtf

Page 4

Db	7756	GTTCATGATCTCAGTAAGTAGGGCTTCAAGAAATCAAGAAATGGCTCAGAAATATTC	7815
Qy	661	AATCTCAATCAAGCGGTATCAGAG	685
Db	7816	AATCTCAATCAAGCGGTATCAGAG	7840

```

RESULT 5
US-08-477-451-5/C
: Sequence 5, Application US/08477451
: Patent No. 5928865
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter Pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335_002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10239 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-477-451-5

```

Query Match	100.0%	Score 685;	DB 2;	Length 10299;
Best Local Similarity	100.0%;	Fred. No. 2.8e-158;		
Matches 685;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	AAAAATGCAAAATATAGATTTTCAGCAAGTATCGCAAGCAAAAAAGCACTTGAAT	60	
Dp	3144	AAAAATGCAAAATATAGATTTTCAGCAAGTATCGCAAGCAAAAAAGCACTTGAAT	3085	
Qy	61	TCGGTAAAGTGTATCATCAATCAAAAGTATCGATTAAGTATATCAATCA	120	
Dp	3084	TCGGTAAAGTGTATCATCAATCAAAAGTATCGATTAAGTATATCAATCA	3025	
Qy	121	GGCGTATCGGTAAAGCAACGGGTGATTCAGTATGGGTAGACAAAGCTTAAGCCAT	180	
Dp	3024	GGCGTATCGGTAAAGCAACGGGTGATTCAGTATGGGTAGACAAAGCTTAAGCCAT	2965	
Qy	181	CTCAAAAAATTTCTCAAGAGCAATGGGCCCAACAGCTCAAAAAATGAAAGTTCAT	240	
Dp	2964	CTCAAAAAATTTCTCAAGAGCAATGGGCCCAACAGCTCAAAAAATGAAAGTTCAT	2905	
Qy	241	GCTGAAAAAAATCTGAAATATATCAATCCGTTAAGATGGTGCAATGGAACCTTAGTC	300	
Dp	2904	GCTGAAAAAAATCTGAAATATATCAATCCGTTAAGATGGTGCAACCTTAGTC	2845	
Qy	301	GCTAATGGTATATCAAGAGCAACCTTTCTAAAACTTTTGACATCAAG	360	

Db 2844 GGTAAAGGGTATCTCAAGCAGAAACCAACAACCTTTCTAAAAAAGCTTTGGCAATCAAG 2785

OY 361 AAAGAGTTGAATGCMAAACTTGGAAATTTCAATTAACATATACATTAATGACATCTCAAAAAC 420

Db 2786 AAAGAGTTGAATGCMAAACTTGGAAATTTCAATTAACATTAATGACATCTCAAAAAC 2725

OY 421 GAACCACTTTATGCTAAAGTTATATTAATAAAGAAAGCAGGCGAAGCAGATCAGCTTGAAGA 480

Db 2724 GAACCACTTTATGCTAAAGTTATATTAATAAAGAAAGCAGGCGAAGCAGATCAGCTTGAAGA 2665

OY 481 CCCATTATAGCTCAAGTGTCTTAATAAAGTAATTCAGAAATTCAGACCAATCAATCAATA 540

Db 2664 CCCATTATAGCTCAAGTGTCTTAATAAAGTAATTCAGAAATTCAGACCAATCAATCAATA 2605

OY 541 GCAGAGTGTTGGTGCTGTGTAAGGCGAAGCAGCGGCTCTCCCTTGAAGAAAGGCATGATTA 600

Db 2604 GCAGAGTGTTGGTGCTGTGTAAGGCGAAGCAGCGGCTCTCCCTTGAAGAAAGGCATGATTA 2545

OY 601 GTTATGATGTCACGTAAAGTGGGCTTTCAAGAGATCAAGAAATGGCTCAGAAATTTGAC 660

Db 2544 GTTATGATGTCACGTAAAGTGGGCTTTCAAGAGATCAAGAAATGGCTCAGAAATTTGAC 2485

OY 661 AATTCATCAAGCCGATTCAGAG 685

Db 2484 AATTCATCAAGCCGATTCAGAG 2460

```

RESULT 6
US-08-477-451-25
/ Sequence 25, Application US/08477451
/ Patent No. 592865
/
GENERAL INFORMATION:
/ APPLICANT: COCCACI, Antonello
/ NUMBER OF SEQUENCES: 46
/ TITLE OF INVENTION: Helicobacter Pylori CagI Region
/
CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Chiron Corporation
/ STREET: 4560 Horton Street
/ CITY: Emeryville
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94608-2916
/
COMPUTER READABLE FORM:
/ MEDIUM TYPE: floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/
CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/477,451
/ FILING DATE: 07-JUN-1995
/
CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION: 435
/ NAME: McCullung, Barbara G.
/ REGISTRATION NUMBER: 33,113
/ REFERENCE/DOCKET NUMBER: 0355_002
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 510-601-2708
/ TELEFAX: 510-655-3542
/
INFORMATION FOR SEQ ID NO: 25:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 19932 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/
MOLECULE TYPE: DNA (genomic)
/
US-08-477-451-25

```

```

Query Match      100.0%; Score 685; DB 2; Length 19923;
Best Local Similarity 100.0%; Pxd. NO. 3.3e-158;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAAAAAGCAAAATTAAGCATTTTCAGCAGGTAAACGCAAAACCACTTGAAAT 60
|||||

```

```

Db 16789 AAAAATGGCAAAATAAGATTTCAGACAGGTACGCAAGCAAAAGCACTTGAAT 16848
Qy 61 TCCGTTAAAGATGATCATCATCAATCAAAAGGTACGATTAAGTTGATATCTCAATCA 120
Db 16849 TCCGTTAAAGATGATCATCATCAATCAAAAGGTACGATTAAGTTGATATCTCAATCA 16908
Qy 121 GCGGTATCACTGCTAAAGCAACGGGTGATTTGATAGGTAGAGCAAGCGTTAGCCGAT 180
Db 16909 GCGGTATCACTGCTAAAGCAACGGGTGATTTGATAGGTAGAGCAAGCGTTAGCCGAT 16968
Qy 181 CTCAAAAATTTCTCAAAAGGCAATTCGCCCAACAGCTCAAAAAATGAAGTCTCAAT 240
Db 16969 CTCAAAAATTTCTCAAAAGGCAATTCGCCCAACAGCTCAAAAAATGAAGTCTCAAT 17028
Qy 241 GCTAGAAAAAATCTGAATATATCAATCGTTAAGATGCTGAATGGAATCCCTAGTC 300
Db 17029 GCTAGAAAAAATCTGAATATATCAATCGTTAAGATGCTGAATGGAATCCCTAGTC 17088
Qy 301 GGTAAATGGTTATCTCAAGCAAGCAACCTCTTCTAAAAAATTTTGGACATCAAG 360
Db 17089 GGTAAATGGTTATCTCAAGCAAGCAACCTCTTCTAAAAAATTTTGGACATCAAG 17148
Qy 361 AAAGATGTAATGCCAAACTTGGAATTTCAATPAACAATAATGACTCAAAAAC 420
Db 17149 AAAGATGTAATGCCAAACTTGGAATTTCAATPAACAATAATGACTCAAAAAC 17208
Qy 421 GAACCAATTAAGTAAAGTTAATAAAGCAAGCGGCAACAGCTAGCTTGAAGA 480
Db 17209 GAACCAATTAAGTAAAGTTAATAAAGCAAGCGGCAACAGCTAGCTTGAAGA 17268
Qy 481 CCCATTTAGCTCAAGTGTCTAAAGATGCAAAATGACCAAGCTCAATCAATA 540
Db 17269 CCCATTTAGCTCAAGTGTCTAAAGATGCAAAATGACCAAGCTCAATCAATA 17328
Qy 541 GCAAGTGTGTTGGGTGTGTAAGGCAAGCGGCTCTTCTTGAAGAAAGCATGATAA 600
Db 17329 GCAAGTGTGTTGGGTGTGTAAGGCAAGCGGCTCTTCTTGAAGAAAGCATGATAA 17388
Qy 601 GTTATGATCTCAGTAAGGAGGCTTTCAAGATCAAGAAATTTGCTCAAGAAATTTGAC 660
Db 17389 GTTATGATCTCAGTAAGGAGGCTTTCAAGATCAAGAAATTTGCTCAAGAAATTTGAC 17448
Qy 661 AATCTCAATCAAGCGGTATCAAGAG 685
Db 17449 AATCTCAATCAAGCGGTATCAAGAG 17473

```

RESULT 7
US-08-993-010-3
Sequence 3, Application US/08993010
Patent No. 6902903

GENERAL INFORMATION:

APPLICANT: Quan, Stella
APPLICANT: Valenzuela, Pablo
TITLE OF INVENTION: HELICOBACTER PYLORI DIAGNOSTICS
FILE REFERENCE: 1360 002
CURRENT FILING DATE: 1997-12-18
EARLIER APPLICATION NUMBER: 60/033,707
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 815
TYPE: DNA
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(804)
US-08-993-010-3

Query Match

82.7%; Score 566.6; DB 3; Length 815;

```

Best Local Similarity 94.2%; Pred. No. 1.5e-129;
Matches 613; Conservative 0; Mismatches 29; Indels 9; Gaps 2;
Qy 1 AAAAATGGCAAAATAAGATTTCAGACAGGTACGCAAGCAAAAGCACTTGAAT 60
Db 10 AAAAATGGCAAAATAAGATTTCAGACAGGTACGCAAGCAAAAGCACTTGAAT 69
Qy 61 TCCGTTAAAGATGATCATCATCAATCAAAAGGTACGATTAAGTTGATATCTCAATCA 120
Db 70 TCCGTTAAAGATGATCATCATCAATCAAAAGGTACGATTAAGTTGATATCTCAATCA 129
Qy 121 GCGGTATCACTGCTAAAGCAACGGGTGATTTGATAGGTAGAGCAAGCGTTAGCCGAT 180
Db 130 GCGGTATCACTGCTAAAGCAACGGGTGATTTGATAGGTAGAGCAAGCGTTAGCCGAT 189
Qy 181 CTCAAAAATTTCTCAAAAGGCAATTCGCCCAACAGCTCAAAAAATGAAGTCTCAAT 240
Db 190 CTCAAAAATTTCTCAAAAGGCAATTCGCCCAACAGCTCAAAAAATGAAGTCTCAAT 249
Qy 241 GCTAGAAAAAATCTGAATATATCAATCGTTAAGATGCTGAATGGAATCCCTAGTC 300
Db 250 GCTAGAAAAAATCTGAATATATCAATCGTTAAGATGCTGAATGGAATCCCTAGTC 309
Qy 301 GGTAAATGGTTATCTCAAGCAAGCAACCTCTTCTAAAAAATTTTGGACATCAAG 360
Db 310 GGTAAATGGTTATCTCAAGCAAGCAACCTCTTCTAAAAAATTTTGGACATCAAG 369
Qy 361 AAAGATGTAATGCCAAACTTGGAATTTCAATPAACAATAATGACTCAAAAAC 417
Db 370 AAAGATGTAATGCCAAACTTTTGAATTTCAATPAACAATAATGACTCAAAAAC 429
Qy 418 AAC-----GAACCAATTAAGTAAAGTTAATAAAGCAAGCGGCAACAGCTAGC 471
Db 430 AACAGCAAGAACCAATTAAGTAAAGTTAATAAAGCAAGCGGCAACAGCTAGC 489
Qy 472 CTTGAAGAACCCATTTAGCTCAAGTGTCTAAAGATGCAAAATTTGACGACTC 531
Db 490 CTTGAAGAACCCATTTAGCTCAAGTGTCTAAAGATGCAAAATTTGACGACTC 549
Qy 532 AATCAATTAAGCAAGTGTGTTGGGTGTGTAAGGCAAGCGGCTCTTCTTGAAGAG 591
Db 550 AATCAATTAAGCAAGTGTGTTGGGTGTGTAAGGCAAGCGGCTCTTCTTGAAGAG 609
Qy 592 CATGATTAAGTTGATCATCATCAATCAAAAGGTACGATTAAGTTGATCATCAATCA 642
Db 610 CATGATTAAGTTGATCATCATCAATCAAAAGGTACGATTAAGTTGATCATCAATCA 660

```

RESULT 8
US-08-053-614-3
Sequence 3, Application US/08053614
Patent No. 5403924

GENERAL INFORMATION:

APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: NEEDLE & ROSENBERG, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/053,614

Tue Mar 7 12:58:26 2006

us-09-360-685c-26.rtf

Page 6

```

? FILING DATE: 19930426
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: SPRATT, GWENDOLYN D.
? REGISTRATION NUMBER: 36 016
? REFERENCE/DOCKET NUMBER: 2200. 009
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 404/688-0770
? TELEFAX: 404/688-9880
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4821 base pairs
? TYPE: NUCLEIC ACID
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1072..4614
? OS-08-053-614-3

```

Query Match	70.8%	Score 485.2	DB 2	Length 4021
Best Local Similarity	86.6%	Pred. No. 2.2e-109		
Matches 561	Conservative 0	Mid. Matches 78	Indels 9	Gaps 2
Qy	1	AAAAATGCAAAAATTAAGATTTTACCAAGGTAACGACAGCAAAAAACGACCTTGAATAT	60	
Db	3319	AAAAATGCAAAAATTAAGATTTTACCAAGGTAACGACAGCAAAAAACGACCAAGAAAT	3378	
Qy	61	TCGGTAAAGATGTATCATCAATCAAAAGTAACGATTAAGTTGATATCTCAATCA	120	
Db	3379	TCGATTAAGATGTATCATCAATCAAAAGTAACGATTAAGTTGATATCTCAATCA	3438	
Qy	121	GGGATTAACATGCTTAAGCAACGGGTATTCATGAGGGTATGAGCAACGTTAGCCGAT	180	
Db	3439	GGGATTAACATGCTTAAGTAAGGGTATTCATGAGGGTATGAGCAACGTTAGCCGAT	3498	
Qy	181	CTCAAAAATTTTCAAAGAGCAATGGCCCAACAGCTCAAAAATAATGAAGTCTCAT	240	
Db	3499	CTCAAAAATTTTCAAAGAGCAATGGCTCAACAGCTCAAAAATAATGAAGTTTCAT	3558	
Qy	241	GCTGAAAAAAATCTGAAATATATCATCCGTTAAGATTGTGTGATGAAACCTTAGTC	300	
Db	3559	GTTGG---AAATCTGAAATATATCATCCGTTAAGATTGTGTGATGAAACCTTAGTC	3615	
Qy	301	GGTATGGGTATCTCAAGACAGACCACTCTTTCTAAAACTTTTGCGACATCAAG	360	
Db	3616	GGTATGGGTATCTGAAATAGAGGCAACGCTCTGCAAAAATTTTTCGAAATCAAG	3675	
Qy	361	AAAGATGTAAGCAAACTGTGAAATTCATATACATTAACATATAGACTCAAAAC	420	
Db	3676	AAAGATGTAAGCAAAATTTTAAAAATTTCAATTAACATTAAGTCTCAAAAC	3735	
Qy	421	-----GAACCAATTTATGCTAAAGTTATATAAAGAAACAGGCGACAGCTAGCCTT	474	
Db	3736	GGCGAGAACCAATTTATGCTCAAGTTATATAAAGAAACAGGCAAGTAGCTAGCCCT	3795	
Qy	475	GAGAACCAATTAACGCTCAAGTTGCTAAAAAGGTAAATGCAAAATATGACCAACTAT	534	
Db	3796	GAGAACCAATTTATGCTCAAGTTGCTAAAAAGGTAACTTAAAAATATGACCAACTAT	3855	
Qy	535	CAAAATGCAAGTGTGGGTGTGTGAGGCAACGACGGGCTCTCCCTTGAAAAGGAT	594	
Db	3856	CAAGACGCAAGATGTTTGGGTGTGTGAGGCAACCGGGGCTCTCCCTTGAAAAGGAT	3915	
Qy	595	GATTAAGTTGATCTCATGTAAGTATGGGCTTTCAAGAAATCAAGAA	642	
Db	3916	GATTAAGTTGATCTCATGTAAGTATGGGCAATCAAGTATGACCTCGAA	3963	

RESULT 9
US-08-316-397B-3
; Sequence 3, Application US/08316397B

1 Patent No.5733740
2 GENERAL INFORMATION:
3 APPLICANT: COVER, TIMOTHY L.
4 APPLICANT: BLASER, MARTIN J.
5 APPLICANT: HARRY KLEANTHOS
6 APPLICANT: TIMOTHY, MORALI K.R.
7 TITLE OF INVENTION: THE TGA GEN AND METHODS FOR DETECTING
8 TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
9 NUMBER OF SEQUENCES: 4
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: NEEDLE & ROSENBERG, P.C.
12 STREET: 1127 Peachtree Street, Suite 1200
13 CITY: Atlanta
14 STATE: Georgia
15 COUNTRY: USA
16 ZIP: 30303
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: floppy disk
19 COMPUTER: IBM PC compatible
20 OPERATING SYSTEM: PC-DOS/MS-DOS
21 SOFTWARE: Patent in Release #1.0, Version #1.25
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/08/316,397B
24 FILING DATE:
25 CLASSIFICATION: 435
26 ATTORNEY/AGENT INFORMATION:
27 NAME: SPRATT, GWENDOLYN D.
28 REGISTRATION NUMBER: 36,016
29 REFERENCE/DOCKET NUMBER: 2200.030
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: 404/688-0770
32 TELEFAX: 404/688-9880
33 INFORMATION FOR SEQ ID NO: 3:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 4821 base pairs
36 TYPE: nucleic acid
37 STRANDEDNESS: double
38 TOPOLOGY: linear
39 MOLECULE TYPE: DNA (genomic)
40 FEATURE:
41 NAME/KEY: CDS
42 LOCATION: 1072..4614
43 IS-08-316-397B-3

Query Match	70.8%	Score 485.2;	DB 2;	Length 4821;
Pred Local Similarity	86.6%;	Pred. No. 2;e-109;		
Matches 561;	Conservative 0;	Mismatches 178;	Indels 9;	Gaps 2

QY	1	AAAAATGGCAAAATAAAGATTTCACGACAGGTAAACGCACAGCAAAAACCGACTTGAAAT	60
Db	3313	AAAAATGGCAAAATAAAGATTTCACGACAGGTAAACGCACAGCAAAAACCGACAAGAAGAAAT	3376
QY	61	TCCGTTAAAGATGTGAATCATCAATCAAAGGTAAACGGATTAAGTGTATATCTCAATCAA	120
Db	3379	TCCATTAAAGATGTGAATCATCAATCAAAGGTAAACGGATTAAGTGTATATCAATCAA	3433
QY	121	GGCGTATCAGTGGCTAAAGCAACGGGTGATTTTCAGTAGGAGTAAAGCAAGCTTACGCCAT	180
Db	3439	GGCGTATCAGTGGCTAAAGCAACGGGTGATTTTCAGTAGGAGTAAAGCAAGCTTACGCCAT	3498
QY	181	CTCAAAATTTCTAAAGAGCAATTGGCCCAACAAGCTCAAAAAAAAAAAGAAAGTCTCAAT	240
Db	3499	CTCAAAATTTCTAAAGAGCAATTGGCTCAACAAGCTCAAAAAAAAAAAGAAAGTTTCAT	3558
QY	241	GCTAAGAAAAAATGTGAATATATCAATCCGTTAAGATGTGTGAATGAAGACCCTAGTGC	300
Db	3559	GTTGG---AAATCTBAATATATCAATCCGTTAAGATGTGTGAAGAGGAACCTAGTGC	3615
QY	301	GGTAATGGGTATATCTCAACGAAGCAACAATCTTTCTTAAAACTTTTGGCATCAAG	360
Db	3616	GGTAATGGGTATATCTGSATTAAGAGCCACAGCTCTGCCAAAAATTTTTTGGATNTCAAG	3675
QY	361	AAGAGTTGAATGCAAAACTTGCAAAATTTCAATTAACAATTAATGCACTCAAAAAC	420

Db 3676 AAGATTGAATGAGAAATTTAAATTTCAATTAACATTAATGCTGCTCAAAAC 3735
QY 421 -----GAACCATTTATGCTAACTATTAATAAGAAAGAGGAGACAGTACCTT 474
Db 3736 GCGGAGAACCCATTTATGCTCAAGTTAATAAGAAACAGGACAGTACCTGACCT 3795
QY 475 GAAGAACCATTTATGCTCAAGTGTCTAAATAAGGTAAATGACCACTCAAT 534
Db 3796 GAAGAACCATTTATGCTCAAGTGTCTAAATAAGGTAAATGACCACTCAAT 3855
QY 535 CAATAGCAAGTGTGTTGGGTGTGTAGAGGCAAGCAGGCTTCCTTTGAAAAGCAT 594
Db 3856 CAGACAGCAAGTGTGTTGGGTGTGTAGAGGCAAGCAGGCTTCCTTTGAAAAGCAT 3915
QY 595 GATAAGTTGATGATCTGATAGTAAAGGCTTTCAAGATCAAGAA 642
Db 3916 GATAAGTTGATGATCTGATAGTAAAGGCGATCAGTTAGCCCTGAA 3963

RESULT 10

US-09-034-306-3
Sequence 3, Application US/09034306
Patent No. 5876943
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: HARRY KLEANTHOS
APPLICANT: TIMMORU, MORALI K.R.
TITLE OF INVENTION: THE TGA GENE AND METHODS FOR DETECTING
NUMBER OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,306
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4821 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1072..4614
US-09-034-306-3

Query Match 70.8%; Score 485.2; DB 2; Length 4821;
Best Local Similarity 86.6%; Pred. No. 2.2e-109;
Matches 561; Conservative 0; Mismatches 78; Indels 9; Gaps 2;

QY 1 AAAATGGCAAAATTAAGATTTCAGCAGTAAACGCAAGAAAGGACCTTGAAAT 60
Db 3319 AAAATGGCAAAATTAAGATTTCAGCAGTAAACGCAAGAAAGGACCTTGAAAT 3378
QY 61 TCCGTTAAAGATGTATCATCAATCAAAAGGTAACGGTAAAGTTGAATCAATCA 120
Db 3379 TCCATTAAGATGTATCATCAATCAAAAGGTAACGGTAAAGTTGAATCAATCA 3438
QY 121 GCGGTATCAGTGTCTAAAGCAAGGCTGATTTCAGTAGGTAGAGCAAGCTTAGCCGAT 180
Db 3439 GCGGTATCAGTGTCTAAAGCAAGGCTGATTTCAGTAGGTAGAGCAAGCTTAGCCGAT 3498
QY 181 CTCAAAAATTTCTCAAGAGCAATTTGCCCCAACAAGCTCAAAAAATGAAGTCTCAAT 240
Db 3499 CTCAAAAATTTCTCAAGAGCAATTTGCCCCAACAAGCTCAAAAAATGAAGTCTCAAT 3558
QY 241 GCTGAAAAAATCTGAAATATATCATCCGTTAAGAAATGCTGAATGAAACCTTACTC 300
Db 3559 GTTG--AAATCTGAAATATATCATCCGTTAAGAAATGCTGAATGAAACCTTACTC 3615
QY 301 GGTATGGTTATCTCAAGCAAGCAACAATCTTTCTAAAACTTTTGGACATCAAG 360
Db 3616 GGTATGGTTATCTGAAATAGAGCCACAGCTCTCCGCAAAATTTTGGATATCAAG 3675
QY 361 AAAGATTGAATGCAAACTTGAAATTTCAATTAACAATTAATGAATCAAAAC 420
Db 3676 AAAGATTGAATGCAAACTTGAAATTTCAATTAACAATTAATGAATCAAAAC 3735
QY 421 -----GAACCATTTATGCTAAAGTTAATAAGAAAGAGGCAAGAGCTAGCTT 474
Db 3736 GCGGAGAACCATTTATGCTCAAGTTAATAAGAAAGAGGCAAGAGCTAGCTT 3795
QY 475 GAAGAACCATTTATGCTCAAGTGTCTAAATAAGGTAAATGACCACTCAAT 534
Db 3796 GAAGAACCATTTATGCTCAAGTGTCTAAATAAGGTAAATGACCACTCAAT 3855
QY 535 CAATAGCAAGTGTGTTGGGTGTGTAGAGGCAAGCAGGCTTCCTTTGAAAAGCAT 594
Db 3856 CAGACAGCAAGTGTGTTGGGTGTGTAGAGGCAAGCAGGCTTCCTTTGAAAAGCAT 3915
QY 595 GATAAGTTGATGATCTGATAGTAAAGGCTTTCAAGATCAAGAA 642
Db 3916 GATAAGTTGATGATCTGATAGTAAAGGCGATCAGTTAGCCCTGAA 3963

RESULT 11

US-09-259-437-3
Sequence 3, Application US/09259437
Patent No. 6153390
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: HARRY KLEANTHOS
APPLICANT: TIMMORU, MORALI K.R.
TITLE OF INVENTION: THE TGA GENE AND METHODS FOR DETECTING
NUMBER OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/259,437
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO:
SEQUENCE CHARACTERISTICS:
LENGTH: 4821 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1072..4614
US-09-259-437-3

Query Match
Best Local Similarity 70.8%; Score 485.2; DB 3; Length 4821;
Matches 561; Conservative 0; Mismatches 78; Indels 9; Gaps 2;

QY 1 AAAAAAGCAAAATTAAGATTTTCAGCAAGTAACGACGACCAAAAGGACCTTGAAT 60
DB 3319 AAAAAAGCAAAATTAAGATTTTCAGCAAGTAACGACGACCAAAAGGACCAAAAGGACCTTGAAT 60
QY 61 TCCGTTAAAGATGTATCATCAATCAAAAGTAACGATTAAGTTGATTAATCTCA 120
DB 3379 TCATTAAGATGTATCATCAATCAAAAGTAACGATTAAGTTGATTAATCTCA 120
QY 121 GCGGTATCAGTGTCTAAAGCAACGAGTATTCAGTACGAGTACCAAGCTTACCGAT 180
DB 3439 GCGGTATCAGTGTCTAAAGCAACGAGTATTCAGTACGAGTACCAAGCTTACCGAT 180
QY 181 CTCAAAAATTTCTCAAGAGCAATTCGACCAAGCTCAAAAGCTCAAAAGCTCAAT 240
DB 3499 CTCAAAAATTTCTCAAGAGCAATTCGACCAAGCTCAAAAGCTCAAAAGCTCAAT 240
QY 241 GCTAGAAAAATTCGAAATATATCATCGTTAAGATGTGAATGGAACCTTAC 300
DB 3559 GTTGG---AAATCTGAATATATCATCGTTAAGATGTGAATGGAACCTTAC 300
QY 301 GGTATGGGTATCTCAAGCAAGCAACCTTCTTAAACCTTTTGGACATCAAG 360
DB 3616 GGTATGGGTATCTCAAGCAAGCAACCTTCTTAAACCTTTTGGACATCAAG 360
QY 361 AAGAGTTGAATGCAAACTGTGAATTTCAATTAACATTAATTAATGACTCAAAAC 420
DB 3676 AAGAGTTGAATGCAAACTGTGAATTTCAATTAACATTAATTAATGACTCAAAAC 420
QY 421 -----GAACCATTTATGCTAAAGTTAATTAAGAAAGAGGCAAGCTTACCTT 474
DB 3736 GCGGAGAGCAACCTTATGCTCAAGTTAATTAAGAAAGAGGCAAGCTTACCTT 474
QY 475 GAAGAACCAATTAAGCTCAAGTTGCTAAAGAAAGTAATTAATTAATGACTCA 534
DB 3796 GAAGAACCAATTAAGCTCAAGTTGCTAAAGAAAGTAATTAATTAATGACTCA 534
QY 535 CAATTAAGCAAGTGTGTTGGGTGTTGAAGCAAGCAAGGAGCTTCCCTTGAAGAGCAT 3855
DB 3856 CAAGCAAGCAAGTGTGTTGGGTGTTGAAGCAAGCAAGGAGCTTCCCTTGAAGAGCAT 3855
QY 595 GATAAAGTTGATGATCTCAAGCAAGCAACCTTCTTAAACCTTTTGGACATCAAG 642
DB 3916 GATAAAGTTGATGATCTCAAGCAAGCAACCTTCTTAAACCTTTTGGACATCAAG 642
RESULT 12

PCT-US93-09782-3
Sequence 3, Application PC/TUS9309782
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: TUMURU, MURALI K. R.
TITLE OF INVENTION: THE LAGA GENE AND METHODS FOR DETECTING
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P. C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/09782
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO:
SEQUENCE CHARACTERISTICS:
LENGTH: 4821 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1072..4614
PCT-US93-09782-3

Query Match
Best Local Similarity 70.8%; Score 485.2; DB 6; Length 4821;
Matches 561; Conservative 0; Mismatches 78; Indels 9; Gaps 2;

QY 1 AAAAAAGCAAAATTAAGATTTTCAGCAAGTAACGACGACCAAAAGGACCTTGAAT 60
DB 3319 AAAAAAGCAAAATTAAGATTTTCAGCAAGTAACGACGACCAAAAGGACCAAAAGGACCTTGAAT 60
QY 61 TCCGTTAAAGATGTATCATCAATCAAAAGTAACGATTAAGTTGATTAATCTCA 120
DB 3379 TCATTAAGATGTATCATCAATCAAAAGTAACGATTAAGTTGATTAATCTCA 120
QY 121 GCGGTATCAGTGTCTAAAGCAACGAGTATTCAGTACGAGTACCAAGCTTACCGAT 180
DB 3439 GCGGTATCAGTGTCTAAAGCAACGAGTATTCAGTACGAGTACCAAGCTTACCGAT 180
QY 181 CTCAAAAATTTCTCAAGAGCAATTCGACCAAGCTCAAAAGCTCAAAAGCTCAAT 240
DB 3499 CTCAAAAATTTCTCAAGAGCAATTCGACCAAGCTCAAAAGCTCAAAAGCTCAAT 240
QY 241 GCTAGAAAAATTCGAAATATATCATCGTTAAGATGTGAATGGAACCTTAC 300
DB 3559 GTTGG---AAATCTGAATATATCATCGTTAAGATGTGAATGGAACCTTAC 300
QY 301 GGTATGGGTATCTCAAGCAAGCAACCTTCTTAAACCTTTTGGACATCAAG 360
DB 3616 GGTATGGGTATCTCAAGCAAGCAACCTTCTTAAACCTTTTGGACATCAAG 360
QY 361 AAGAGTTGAATGCAAACTGTGAATTTCAATTAACATTAATTAATGACTCAAAAC 420

Db 3676 AAGAAATGATGAGAAATTTAAATTTCAATACATACATATATGCTCAAAAAC 3735
Qy 421 -----GACCCTTTATGCTAAAGTAAATTAAGAAAGACAGGCGACAGCTAGCCTT 474
Db 3736 GCGGAGAACCCCTTTATGCTCAAGTTAAATTAAGAAACAGACAGTAGCTAGCCCTT 3795
Qy 475 GAAGAACCATTTACCTCAAGTTGCTAAATAGTAATGCAAAATATGACCGACTCAAT 534
Db 3796 GAAGAACCATTTATGCTCAAGTTGCTAAATAGTAATGCAAAATATGACCACTCAAT 3855
Qy 535 CAATAGCAAGTGTGGTGTGTGTAGGCAAGCAGCGGCTTCCCTTTGAAGAGCAT 594
Db 3856 CAAGCAGCAGCAAGTGTGGTGTGTGTAGGCAAGCAGCGGCTTCCCTTTGAAGAGCAT 3915
Qy 595 GATTAAGTATGATCTCAGTAAGGTAAGGCTTTCAGAGATCAAGA 642
Db 3916 GATTAAGTATGATCTCAGTAAGGTAAGGCTTTCAGAGATCAAGA 3963

RESULT 13

US-08-053-614-1
Sequence 1, Application US/08053614
Patent No. 5403924
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: TOMMORU, MURALI K.R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/053,614
FILING DATE: 19930426
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-9880
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3648 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: CDS
LOCATION: 1072..3648
US-08-053-614-1

Query Match 40.5%; Score 277.2; DB 2; Length 3648;
Best Local Similarity 92.1%; Pred.No. 2e-58;
Matches 304; Conservative 0; Mismatches 23; Indels 3; Gaps 1;
Qy 1 AAAAAAGCAAAATATAGATTTTCAGCAAGTATACGCAAGCAAAAGCAAGCACTTGAAT 60

Db 3319 AAAAAAGCAAAATATAGATTTTCAGCAAGTATACGCAAGCAAAAGCAAGCAAGCAAT 3378
Qy 61 TCCGTTAAAGATGTATCATCATCAATCAAAAGTAAAGTAAAGTATATCTCAATCA 120
Db 3379 TCCATTAAGATGTATCATCATCAATCAAAAGTAAAGTAAAGTATATCTCAATCA 3438
Qy 121 GCGGTATCAGTGGCTTAAGCAAGGGTATTTCAAGTGGGTAGAGCAAGCTTAGCCGAT 180
Db 3439 GCGGTATCAGTGGCTTAAGTATACGTCGATTTTCAGTGGGTAGAGCAAGCTTAGCCGAT 3498
Qy 181 CTCAAAAATTTCTCAAGAGCAATTTGGCCCAAGCTCAAAAAATGAAGTCAAT 240
Db 3499 CTCAAAAATTTCTCAAGAGCAATTTGGCTCAACAAAGCTCAAAAAATGAAGTCAAT 3558
Qy 241 GCTAGAAAAATCTGAATATATCATCCGTTAAGATGTGTGAATGAAGCAAGCTAGTC 300
Db 3559 GTTGG--AAATTCGAATATATCAATCCGTTAAGATGTGTGAAGCAAGCTAGTC 3615
Qy 301 GGTAAATGGTTATCTCAAGCAGAGCCACA 330
Db 3616 GGTAAATGGTTATCTCAAGATAGGCCACA 3645

RESULT 14

US-08-316-397B-1
Sequence 1, Application US/08316397B
Patent No. 5733740
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: HARRY KLEANTHOUS
APPLICANT: TOMMORU, MURALI K.R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-9880
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3648 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: CDS
LOCATION: 1072..3648
US-08-316-397B-1

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2006, 21:25:19 ; Search time 555 Seconds

(without alignments)
8225.792 Million cell updates/sec

Title: US-09-360-685C-26

Sequence: 1 aaaatggcaaaataagga.....caatcaagcgatcagaag 685

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Maximum Match 0%

Listing first 45 summaries

Database :
1: N_Geneseq_21.*
2: geneseqn1908.*
3: geneseqn1908.*
4: geneseqn2000.*
5: geneseqn2001.*
6: geneseqn2002.*
7: geneseqn2003.*
8: geneseqn2004.*
9: geneseqn2005.*
10: geneseqn2006.*
11: geneseqn2007.*
12: geneseqn2008.*
13: geneseqn2009.*
14: geneseqn2010.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	685	100.0	4149	2	AA62461 Helicobac
2	685	100.0	5925	2	AA62461 Helicobac
3	685	100.0	19932	2	AA62461 Helicobac
4	531.6	77.6	3558	13	ADU05406 DNA encod
5	502.6	73.4	1024	6	ABX66359 Helicobac
6	497.6	72.6	1024	6	ABX66359 Helicobac
7	485.2	70.8	4821	2	AA086728 Helicobac
8	485.2	70.8	4821	2	AA086728 Helicobac
9	485.2	70.8	4821	2	AA086728 Helicobac
10	482	70.4	3546	3	AA691795 Helicobac
11	477.2	69.7	3534	2	AA691795 Helicobac
12	477.2	69.7	3534	2	AA691795 Helicobac
13	477.2	69.7	3534	2	AA691795 Helicobac
14	406.2	59.3	1024	6	ABX66128 Helicobac
15	379.4	55.4	854	6	ABX66698 Helicobac
16	379.4	55.4	1024	6	ABX66370 Helicobac
17	379.4	55.4	1024	6	ABX66370 Helicobac
18	277.2	40.5	3648	2	AA064580 Helicobac
19	277.2	40.5	3648	2	AA064580 Helicobac

20	271.2	39.6	762	6	ABX66257 Helicobac
21	268.4	39.2	1008	6	ABX66257 Helicobac
22	48.8	7.1	1024	6	ABX66874 Helicobac
23	48	7.0	12138	6	ABX66874 Helicobac
24	48	7.0	12138	6	ABX66874 Helicobac
25	47.6	6.9	863	4	AA195036 DNA trans
26	47	6.9	14006	6	ABX66358 Helicobac
27	46.8	6.8	5542	6	ABX66358 Helicobac
28	46.8	6.8	9964	6	ABX66358 Helicobac
29	46.8	6.8	10205	6	ABX66358 Helicobac
30	46.8	6.8	10205	6	ABX66358 Helicobac
31	46.4	6.8	5127	4	AA545425 Helicobac
32	46	6.7	7969	6	ABX667037 Helicobac
33	46	6.7	7969	6	ABX667037 Helicobac
34	45.6	6.7	6182	6	ABX667037 Helicobac
35	45.4	6.6	14362	6	ABX667037 Helicobac
36	45.4	6.6	14708	4	AA545514 Helicobac
37	45.4	6.6	14708	4	AA545514 Helicobac
38	45.4	6.6	14708	6	ABX662325 Helicobac
39	45.4	6.6	14708	6	ABX662325 Helicobac
40	45.2	6.6	8222	8	ACF62794 Helicobac
41	45.2	6.6	8222	8	ACF62794 Helicobac
42	45.2	6.6	11222	10	ADB54190 Helicobac
43	45.2	6.6	11222	10	ADB54190 Helicobac
44	45	6.6	5814	6	ABX663561 Helicobac
45	45	6.6	9666	6	ABX663561 Helicobac

ALIGNMENTS

RESULT 1	AA62461	standard; DNA; 4149 BP.
ID	AA62461	standard; DNA; 4149 BP.
XX	AA62461	
AC	AA62461	
XX	AA62461	
DT	17-OCT-2003	(revised)
DT	15-FEB-1999	(first entry)
XX		
DE	Helicobacter pylori caga-Vibrio cholerae A2+B subunit gene fusion.	
XX		
KW	Vaccine; antigen; CagB; antigen; toxin; diagnosis; gastritis; ulcer;	
KW	stomach cancer; ss.	
XX		
OS	Helicobacter pylori.	
OS	Vibrio cholerae.	
OS	Chimeric.	
XX		
Key	Location/Qualifiers	
FT	1..3444	
FT	/*tag= a	
FT	/product= "Caga"	
FT	3444..4020	
FT	/*tag= b	
FT	/product= "A2+B subunits"	
CDS		
XX		
WO9844130-A1.		
XX		
08-OCT-1998.		
XX		
PF	31-MAR-1998;	98WO-KR000073.
XX		
PR	31-MAR-1997;	97KR-00011950.
XX		
PR	31-MAR-1997;	97KR-00011951.
XX		
PA	(DAEW-) DAEWOONG PHARM CO LTD.	
XX		
PI	Kim B, Shin S, Yu Y, Park M, Choi D, Jung H;	
XX		
DR	WPI, 1998-568279/48.	
XX		
DR	P-PSDB; AAM80600.	
XX		

PT New chimeric proteins for use against *Helicobacter pylori* - comprising an
 PT antigenic protein of *H. pylori* and A1 and B subunits of Vibrio cholerae
 PS toxin, preferably produced by recombinant techniques.
 Claim 4; Page 78-80; 102pp; English.

CC This is the nucleotide sequence of a fusion gene prepared by ligating the
 CC cagA gene of *Helicobacter pylori* and the A2 and B subunits genes of
 CC Vibrio cholerae toxin. It codes for a claimed 138-amino acid fusion
 CC protein (see AAM8600). The invention relates to: chimeric proteins
 CC comprising antigenic proteins of *H. pylori* and A2 and B subunits of *V.*
 CC cholerae toxin; recombinant DNAs encoding such chimeric proteins;
 CC recombinant expression vectors; a process for preparing the chimeric
 CC proteins using transformed microbial host cells; and to preventative and
 CC therapeutic vaccines comprising the chimeric proteins for *H. pylori*-
 CC associated diseases such as gastritis, gastric ulcer, duodenal ulcer and
 CC gastric cancer (claimed). The chimeric proteins are designed to have
 CC excellent immunogenicity, to be stable in the stomach, to penetrate the
 CC mucous membrane of the intestine, and to stimulate production of sIgA.
 CC They can additionally be used as active ingredients in diagnostic kits
 CC for *H. pylori* infection, and for production of anti-*H. pylori* antibody.
 CC (Updated on 17-Oct-2003 to standardise OS field)

Sequence 4149 BP; 1544 A; 714 C; 819 G; 1072 T; 0 U; 0 Other;

Query Match Best Local Similarity 100.0%; Score 685; DB 2; Length 4149;
 Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAGCAAAATTAAGATTTTCAGCAGGTACGACGCAAAAGCACTTGAAT 60
 DB 2248 AAAAAAGCAAAATTAAGATTTTCAGCAGGTACGACGCAAAAGCACTTGAAT 60
 QY 61 TCCGTTAAAGATGTATCATCATCAAAAGGTAAAGCAATTTGATATCTCAATCA 120
 DB 2308 TCCGTTAAAGATGTATCATCATCAAAAGGTAAAGCAATTTGATATCTCAATCA 120
 QY 121 GCGGTATCAGTGGCTAAAGCAAGCGGTATTCAGTGGGTAGACAGCGTTAGCCAT 180
 DB 2368 GCGGTATCAGTGGCTAAAGCAAGCGGTATTCAGTGGGTAGACAGCGTTAGCCAT 180
 QY 181 CTCAAAAATTTCTCAAGAGCAATTTGCCCAACACTCCAAAAATGAAGTTCAT 240
 DB 2428 CTCAAAAATTTCTCAAGAGCAATTTGCCCAACACTCCAAAAATGAAGTTCAT 240
 QY 241 GCTAGAAAAAATTCGAATATATCAATCCGTAAAGATGCTGAATGAACTTATC 300
 DB 2488 GCTAGAAAAAATTCGAATATATCAATCCGTAAAGATGCTGAATGAACTTATC 300
 QY 301 GGTATGAGGTATCTCAAGAGCAAGCACTTTCTTAAAACTTTTGGACATCAAG 360
 DB 2548 GGTATGAGGTATCTCAAGAGCAAGCACTTTCTTAAAACTTTTGGACATCAAG 360
 QY 361 AAGAGTTGAATGCAAACTTGAATTTCAATACATTAACATTAATGAACTCAAAAC 420
 DB 2608 AAGAGTTGAATGCAAACTTGAATTTCAATACATTAACATTAATGAACTCAAAAC 420
 QY 421 GAACCATTTATGCTAAAGTAAATTAAGAAAGCAAGCGGCAAGCACTTGAAGAA 480
 DB 2668 GAACCATTTATGCTAAAGTAAATTAAGAAAGCAAGCGGCAAGCACTTGAAGAA 480
 QY 481 CCCATTTACGCTCAAGTCTTAAAGGTAAATGCAAAATTTGACGCACTCAATCAATA 540
 DB 2728 CCCATTTACGCTCAAGTCTTAAAGGTAAATGCAAAATTTGACGCACTCAATCAATA 540
 QY 541 GCAAGTGTGTTGGGTGTTTGAAGCAAGCAAGCGGCTTCCCTTTGAAAAGGATGATTA 600
 DB 2788 GCAAGTGTGTTGGGTGTTTGAAGCAAGCAAGCGGCTTCCCTTTGAAAAGGATGATTA 600
 QY 601 GTTATGATCTCAGTAAGGTAGGCTTTCAAGGAATCAAGAAATGGCTAGAAAAATTGAC 660
 DB 2848 GTTATGATCTCAGTAAGGTAGGCTTTCAAGGAATCAAGAAATGGCTAGAAAAATTGAC 660

QY 661 AATTCATCAAGCGGTATCAAG 685
 DB 2908 AATTCATCAAGCGGTATCAAG 2932

RESULT 2
 AA048733
 ID AA048733 standard; DNA; 5925 BP.
 AC AA048733;
 XX
 DN 25-MAR-2003 (revised)
 DT 17-MAR-1994 (first entry)
 XX
 DE CAI gene.

XX Cytotoxin; CT; *H. pylori*; precursor; vacuolation; cell death; eukaryote;
 KW cytotoxin-associated immunodominant antigen; CAI heat shock protein;
 KW hsp60; type B; gastritis; peptic ulcer; gastric tumours; ss.
 XX
 OS *Helicobacter pylori*.

XX Key Location/Qualifiers
 FH 525..529
 FT RBS /+cag= b
 FT CDS 535..3978
 FT /+cag= a
 FT /product= "CAI antigen"
 FT terminator 4010..4032
 FT /+cag= c
 FT /standard_name= "Rho_independent_terminator"

XX W09318150-A1.
 XX 16-SEP-1993.
 XX
 XX 02-MAR-1993; 93WO-BP000472.
 XX
 XX 02-MAR-1992; 92IT-FI000052.
 PR 25-JAN-1993; 93WO-BP000158.
 XX
 XX (ISTS) BIOCLINE SCLAVO SPA.
 PA
 XX
 XX
 PI Covacci A, Bugnoli M, Telford J, Macchia G, Rappuoli R;
 DR WPI; 1993-303464/38.
 DR P-PSDB; AAR41199.
 XX

PT Recombinant *Helicobacter pylori* protein and corresp. gene - is a
 PT cytotoxin, antigen or heat shock protein used for treating and preventing
 PT type B gastritis, gastric ulcers and gastric tumours.
 XX
 XX
 PS Claim 28; Fig 4; 83pp; English.

CC This sequence encodes the cytotoxin-associated immunodominant antigen,
 CC CAI, of *H. pylori*. The protein is a hydrophilic, surface-exposed protein
 CC having a molecular weight of approximately 120-132 kD, pI 128-130 kD,
 CC and an isoelectric point of 9.72. *H. pylori* isolates which do not
 CC produce the CAI antigen do not have the *cag* gene, and are also unable to
 CC produce an active cytotoxin. The association between the presence of the
 CC *cag* gene and cytotoxicity suggests that the product of the *cag* gene is
 CC necessary for the transcription, folding, export or function of the
 CC cytotoxin. Alternatively both the *cag* gene and the *ct* gene are absent in
 CC noncytotoxic strains suggesting physical linkage between the genes. The
 CC absence of a typical leader sequence suggests the presence of an
 CC independent export system. Computer searches for promoter regions in the
 CC region upstream from the ATG codon identified sequences resembling either
 CC -10 or -35 regions, however a region with a good consensus to the *E. coli*
 CC or *H. pylori* promoter sequences was not found. The CAI antigen is very
 CC hydrophilic and does not show obvious leader peptide or transmembrane
 CC sequences. The most hydrophilic region is from amino acids 600-900, where
 CC a number of unusual features are observed. There is a repetition of the
 CC sequences EFNGKNKDFSK and EPTIA, and the presence of a stretch of six

CC contiguous asparagines. This protein, and others derived from *H. pylori*,
 CC esp. cytotoxin (Ct) or a heat shock protein (see also AA048732 and
 CC AA048734), may be used to treat, prevent and diagnose *H. pylori*
 CC infection. *H. pylori* is the causative agent of type B gastritis, peptic
 CC ulcers and gastric tumours. (Updated on 25-MAR-2003 to correct PW field.)
 CC (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to
 CC correct PA field.)

XX
 SQ Sequence 5925 BP; 2076 A; 1057 C; 1138 G; 1654 T; 0 U; 0 Other;

Query Match 100.0%; Score 685; DB 2; Length 5925;
 Best Local Similarity 100.0%; Pred. No. 66-159;
 Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAATGGCAAAATATAGATTTCAGCAAGGTACGCAAGCAAAAGCCGCTTGAAT 60
 DB 2782 AAAAATGGCAAAATATAGATTTCAGCAAGGTACGCAAGCAAAAGCCGCTTGAAT 2841
 QY 61 TCCGTTAAAGATGTATCATCATCAAAAGGTAAAGGTAAAGTATATCTCAATCAA 120
 DB 2842 TCCGTTAAAGATGTATCATCATCAAAAGGTAAAGGTAAAGTATATCTCAATCAA 2901
 QY 121 GCGGTATCAGTGTCTAAAGCAAGCGGTATTCAGTAGGGTAGAGCAAGCGTTAGCCGAT 180
 DB 2902 GCGGTATCAGTGTCTAAAGCAAGCGGTATTCAGTAGGGTAGAGCAAGCGTTAGCCGAT 2961
 QY 181 CTCAAAAATTTCTCAAGGAGCAATGGCCCAACAGCTCAAAAAATGAAAGTCTCAT 240
 DB 2962 CTCAAAAATTTCTCAAGGAGCAATGGCCCAACAGCTCAAAAAATGAAAGTCTCAT 3021
 QY 241 GCTAGAAAAAATCTGAAATATATCATCCGTAAAGATGTGTGAATGGAACCTAGTC 300
 DB 3022 GCTAGAAAAAATCTGAAATATATCATCCGTAAAGATGTGTGAATGGAACCTAGTC 3081
 QY 301 GGTATGGTATCTCAAGCAAGGCACAACCTTTCTTAAAACTTTTCGACATCAAG 360
 DB 3082 GGTATGGTATCTCAAGCAAGGCACAACCTTTCTTAAAACTTTTCGACATCAAG 3141
 QY 361 AAAGATTGAATGCAAACTGTGAATTTCAATTAACATTAATGATCTCAAAAC 420
 DB 3142 AAAGATTGAATGCAAACTGTGAATTTCAATTAACATTAATGATCTCAAAAC 3201
 QY 421 GAACCATTTATGCTAAAGTAAATTAAGAAAGAGGCAAGCAGCTAGCCCTGAAGAA 480
 DB 3202 GAACCATTTATGCTAAAGTAAATTAAGAAAGAGGCAAGCAGCTAGCCCTGAAGAA 3261
 QY 481 CCCATTACGCTCAAGTGTCTTAAAGGTAAATGCAAAATTCAGCCGCTCAATCAATA 540
 DB 3262 CCCATTACGCTCAAGTGTCTTAAAGGTAAATGCAAAATTCAGCCGCTCAATCAATA 3321
 QY 541 GGAAGTGTGGT 600
 DB 3322 GGAAGTGTGGT 3381
 QY 601 GTTGATGATCTCAGTAGAGGCTTTCAAGGAATCAGGAATGTGCGTCAAGAAATTTGAC 660
 DB 3382 GTTGATGATCTCAGTAGAGGCTTTCAAGGAATCAGGAATGTGCGTCAAGAAATTTGAC 3441
 QY 661 AATCTCAATCAAGCGGTATCAAGAG 685
 DB 3442 AATCTCAATCAAGCGGTATCAAGAG 3466

RESULT 3
 AAT46159
 ID AAT46159 standard; DNA; 19932 BP.

AC AAT46159;
 XX
 DT 30-JUN-1997 (first entry)
 XX
 DE CagI locus.
 XX

KW CagI; CagA; virulence factor; exporter molecule; homology; pfl gene;
 KW Bordetella pertussis; VIB B; Agrobacterium tumefaciens; invasion factor;
 KW Salmonella; type I strait; virulence; diagnosis; *H. pylori*; infection;
 KW vaccine; treatment; duodenal; gastric ulcer; active gastritis;
 KW adenocarcinoma; sg.
 XX
 OS Helicobacter pylori.

XX	PH	Key	Location/Qualifiers
FT	FT	CDS	389..1579
FT	FT		/*tag= a
FT	FT		/note= "putative open reading frame; no start codon given"
FT	FT	CDS	complement(1778..2173)
FT	FT		/*tag= e
FT	FT		/note= "putative open reading frame; no start codon given"
FT	FT	CDS	1794..2168
FT	FT		/*tag= b
FT	FT		/note= "putative open reading frame; no start codon given"
FT	FT	CDS	complement(2591..3001)
FT	FT		/*tag= f
FT	FT		/note= "putative open reading frame; no start codon given"
FT	FT	CDS	complement(3020..3259)
FT	FT		/*tag= g
FT	FT		/note= "putative open reading frame; no start codon given"
FT	FT	CDS	3343..4491
FT	FT		/*tag= c
FT	FT		/note= "putative open reading frame; no start codon given"
FT	FT	CDS	4488..5426
FT	FT		/*tag= d
FT	FT		/note= "putative open reading frame; no start codon given"
FT	FT	CDS	complement(5508..6233)
FT	FT		/*tag= h
FT	FT		/note= "putative open reading frame; no start codon given"
FT	FT	CDS	complement(6218..6613)
FT	FT		/*tag= h
FT	FT		/note= "putative open reading frame; no start codon given"
FT	FT	CDS	complement(6562..7074)
FT	FT		/*tag= i
FT	FT		/note= "putative open reading frame; no start codon given"
FT	FT	CDS	complement(7007..7375)
FT	FT		/*tag= j
FT	FT		/note= "putative open reading frame; no start codon given"
FT	FT	CDS	complement(7371..7802)
FT	FT		/*tag= k
FT	FT		/note= "putative open reading frame; no start codon given"
FT	FT	CDS	complement(7975..8454)
FT	FT		/*tag= l
FT	FT		/note= "putative open reading frame; no start codon given"
FT	FT	CDS	complement(8496..8918)
FT	FT		/*tag= m
FT	FT		/note= "putative open reading frame; no start codon given"
FT	FT	CDS	complement(8977..9762)
FT	FT		/*tag= n
FT	FT		/note= "putative open reading frame; no start codon given"
FT	FT	CDS	complement(9784..10575)
FT	FT		/*tag= o
FT	FT		/note= "putative open reading frame; no start codon given"

FT	CDS	10557..12758
FT		/tag= "p"
FT		/note= "putative open reading frame; no start codon given"
FT	CDS	complement(12743..13420)
FT		/tag= "q"
FT		/note= "putative open reading frame; no start codon given"
FT	CDS	complement(13374..13742)
FT		/tag= "r"
FT		/note= "putative open reading frame; no start codon given"
FT	CDS	complement(13666..14129)
FT		/tag= "g"
FT		/note= "putative open reading frame; no start codon given"
FT	CDS	complement(19256..19832)
FT		/tag= "c"
FT		/note= "putative open reading frame; no start codon given"

W09633274-A1.

MO9633274-A1.
24-OCT-1996.
18-APR-1996; 96MO-IB000343
20-APR-1995; 95US-00425194
07-JUN-1995; 95US-00477451
(BIOC-) BIOCINE SPA.
Covacci A;
MPI; 1996-485780/48.

Helicobacter pylori CagI polynucleotide and related proteins - used in diagnosis and in vaccines for the treatment of *H. pylori* infection associated disease.

Claim 2; Fig 3A-R; 303pp; English.

The present sequence is the complete nucleotide sequence of the CagI locus, including the true 5'-terminus of CagI. The sequence was constructed using overlapping clones (one of which also overlaps with the CagA region). The CagI region contains clusters of putative open reading frames (ORFs) with different polarities. The putative ORFs for this region are shown above. It is hypothesized that some of these ORFs may encode exporter molecules with homology to the *peI* genes of *Bordetella pertussis* and *VIR B4* genes of *Aerobacterium tumefaciens* and for proteins with motifs shared by the purported invasion factors of *Salmonella* genus. The absence of CagI sequences in the type I strains is associated with the absence of CagI sequences (which may encode virulence factors restricted to type I strains). The CagI nucleotide sequence, its fragments and type I strain infection in an individual and in vaccines (claimed) encoded proteins are used in the diagnosis of *H. pylori* (esp. *H. pylori* type I strain) infection in an individual and in vaccines (claimed) for the treatment of *H. pylori* infection associated with e.g. duodenal and gastric ulcers, severe forms of active gastritis (esp. type gastritis) and gastric adenocarcinoma.

Sequence 19932 BP; 6677 A; 3635 C; 3483 G; 6137 T; 0 U; 0 Other

[illegible]

	0	mismatches	0;	Indels	0;	Gaps	0;
1 AAAAAATGCAAAAATAAGCATTTTTCCTCCACCGAATGGTAAATTA							

Db 16789 AAAAAATGCAAAATAAGGATTTCAGCAGCGTACCGCAAGCAAAAAACGACCTTGAAT 1684

Qy 61 TCCGTTAAAGATGTGATCATCAATCAAAAGTAACGATAAAGTTGATATCTCAATCAA 120

Db 16949 TCCGTTAAGATGTGATCATCTTC 120

CGAATGATATCATCA 1690

QY 121 GGGGATATCATGGCTAAAGCAACGGGTGATTTTCATGAGGTAGACCAACCGTTAGCCGAT 180
 Db 16909 GCGGATATCATGGCTAAAGCAACGGGTGATTTTCATGAGGTAGACCAACCGTTAGCCGAT 16909
 QY 181 CTCAAAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCAAAAAATGAAAGTCTCAAT 240
 Db 16969 CTCAAAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCAAAAAATGAAAGTCTCAAT 17029
 QY 241 GGTGAAAAAATCTGCAATTTATCATCCGTTAGAGTGGTGTAGATGAAACCGTAGTC 300
 Db 17029 GGTGAAAAAATCTGCAATTTATCATCCGTTAGAGTGGTGTAGATGAAACCGTAGTC 17089
 QY 301 GGTAAAGGGTATCTCAAGCAAGGCAACCTCTTCTAAAACTTTTGGACATCAAG 360
 Db 17089 GGTAAAGGGTATCTCAAGCAAGGCAACCTCTTCTAAAACTTTTGGACATCAAG 17149
 QY 361 AAAAGTGTGAATGCAAACTTGGAAATTTCAATTAACAATATATATGATGACTCAAAAC 420
 Db 17149 AAAAGTGTGAATGCAAACTTGGAAATTTCAATTAACAATATATATGATGACTCAAAAC 17209
 QY 421 GAACCCATTATGCTAAAGTTAATTAATAAAGAAAGCAAGGCGACGCTACCTGTGAAGA 480
 Db 17209 GAACCCATTATGCTAAAGTTAATTAATAAAGAAAGCAAGGCGACGCTACCTGTGAAGA 17269
 QY 481 CCCATTATAGCTCAAGTTGCTAAAGATAATGCAAAAAATGACACGACTCATCAATA 540
 Db 17269 CCCATTATAGCTCAAGTTGCTAAAGATAATGCAAAAAATGACACGACTCATCAATA 17329
 QY 541 GCAAGGCTTTGGGTGTGTGAGGCGACAGCGGGCTTCCCTTTGAAAAAGGATGATAA 600
 Db 17329 GCAAGGCTTTGGGTGTGTGAGGCGACAGCGGGCTTCCCTTTGAAAAAGGATGATAA 17389
 QY 601 GTTGATGATCTCAAGTAGAGTGAAGGCTTTCAAGGAATCAAGAAATTTGGCTCAGAAAAATTGAC 660
 Db 17389 GTTGATGATCTCAAGTAGAGTGAAGGCTTTCAAGGAATCAAGAAATTTGGCTCAGAAAAATTGAC 17449
 QY 661 AATTCATCAAGCGGTATCAAGAAG 685
 Db 17449 AATTCATCAAGCGGTATCAAGAAG 17473

RESULT 4
ADU05406
ID ADU05406 standard; DNA; 3558 BP
XX

AC ADU05406
XX

DT 27-JAN-2005 (first entry
XX

DNA encoding H. pylori antigenic protein HP0547

KW ds; gene; antibacterial; antigenic; H. pylori infection.
XX

OS Helicobacter pylori
XX

PN WO2004094467-A2
XX

04-NOV-2004

PF 22-APR-2004
XX

22-APR-2003; 2003EP-00450097

PA (INTE-) INTERCELL AG.
XX

Meinke A, Min Bui D, Nagy E, Henics T,

DR WPI; 2004-775908/76.
DR P-PSDB: ADN0EEA

XX
PT

reactive antigens from *Helicobacter pylori*, and encoding nucleic acid molecules, useful for diagnosing, preventing or

XX (UYVA-) UNITV VANDERBILT.
PA
XX
XX
PI Cover TL, Blaser MJ, Tummuru MKR;
XX
XX WPI; 1994-151235/18.
DR P-PSDB; AAR53269.
XX
XX
PT DNA coding tag A gene, from *Helicobacter pylori* - useful for detecting
PT predisposition to peptic ulceration.
XX
XX
PS Claim 27; Page 63-68; 87pp; English.
XX
XX
XX Monoclonal antibodies directed against the antigen may be used to detect
CC tag A antigen presence which is indicative of a predisposition to peptic
CC ulceration. A ligand e.g. Antibody, specifically reactive with the tag A
CC antigen can be used to treat peptic ulcers. (Updated on 25-MAR-2003 to
XX correct PN field.)

Query Match	Best Local Similarity	Matches	Score	DB	Length
70.8%;	86.6%;	561;	485.2;	4821;	
Conservative	Pred. No. 1.3e-109;	0;	Mismatches 78;	Indels 9;	Gaps 2;
Sequence 4821 BP; 1721 A; 874 C; 930 G; 1296 T; 0 U; 0 Other;					
1	AAAAATGGCAAAATTAAGATTCAGACAGGTAAACGATTAAGTGAATCTCAATCAA	60			
3319	AAAAATGGCAAAATTAAGATTCAGACAGGTAAACGATTAAGTGAATCTCAATCAA	3378			
61	TCCGTTAAGATGATCATCAATCAAAAGGTAAACGATTAAGTGAATCTCAATCAA	120			
3379	TCCATTAAGATGATCATCAATCAAAAGGTAAACGATTAAGTGAATCTCAATCAA	3438			
121	GCGGTATCAGTGGCTTAAAGCAACGGGTGATTCTAGTAGGTTAGCAACGGTTAGCCGAT	180			
3439	GCGGTATCAGTGGCTTAAAGCAACGGGTGATTCTAGTAGGTTAGCAACGGTTAGCCGAT	3498			
181	CTCAAAATTTCTCAAAAGGAGCAATTGGCCCAACAGCTCAAAAATTAAGAAAGTCTCAAT	240			
3499	CTCAAAATTTCTCAAAAGGAGCAATTGGCCCAACAGCTCAAAAATTAAGAAAGTCTCAAT	3558			
241	GCTAAGAAAAAATCTGAATATATATCAATCCGTTAAGATGATGATGAATGAAACCTTAGTC	300			
3559	GTTGG---AAATCTGAATATATATCAATCCGTTAAGATGATGATGAATGAAACCTTAGTC	3615			
301	GCTAATGGGTATCTCAGACGAAAGCCAACTTTCTTAAAACTTTTGGACATCAAG	360			
3616	GCTAATGGGTATCTCAGACGAAAGCCAACTTTCTTAAAACTTTTGGACATCAAG	3675			

QY 361 AAAGTTGATGCAAACTTGAAATTTCAATTAACATTAATGACTCAAAAC 420
 Db 3676 AAAGATTAATGAGAAATTTAAATTTCAATTAACATTAATGACTCAAAAC 3735
 QY 421 -----GAACCATTTATGCTTAAGTTAAATTAAGGAGGCAAGGAGTACCTT 474
 Db 3736 GCGGAGAACCATTTATGCTTAAGTTAAATTAAGGAGGCAAGGAGTACCTT 3795
 QY 475 GAAGAACCATTTATGCTTAAGTTAAATTAAGGAGGCAAGGAGTACCTT 534
 Db 3796 GAAGAACCATTTATGCTTAAGTTAAATTAAGGAGGCAAGGAGTACCTT 3855
 QY 535 CAATTAAGCAATGTTGGGTGTTGTAAGGCAAGGAGGCAAGGAGTACCTT 594
 Db 3856 CAAGCAAGCAAGTGTGTTGTTGTAAGGCAAGGAGGCAAGGAGTACCTT 3915
 QY 595 GATTAAGTTGATGCTCAAGTAAGGAGGCTTCAAGGAGTACCAAGA 642
 Db 3916 GATTAAGTTGATGCTCAAGTAAGGAGGCTTCAAGGAGTACCAAGA 3963

RESULT 9

AAAT14051 standard; DNA; 4821 BP.

AAAT14051;
 AC 09-SEP-1996 (first entry)
 DT Helicobacter pylori TagA gene encoding a 120-128 kD antigen.
 DE Helicobacter pylori TagA gene encoding a 120-128 kD antigen.
 KW Antigen; Peptic ulcer; chronic gastritis; gastric adenocarcinoma;
 XX diagnosis; predilection; antibody; vaccine; infection; ds.
 OS Helicobacter pylori.

PH Key Location/Qualifiers
 FT 5'UTR 1..1071
 FT CDS /*tag= b
 FT 1072..4617
 FT /*tag= a
 FT 4618..4821
 FT /*tag= c

XX W09610639-A2.
 PD 11-APR-1996.
 XX 29-SEP-1995; 95WO-US012669.
 PF 30-SEP-1994; 94US-00316397.
 XX (UYVA-) UNITV VANDERBILT.
 PA (ORAV-) ORAVAX INC.
 XX

PI Cover TL, Blaser MJ, Kleantous H, Tummur MKR;
 DR WPI; 1996-209361/21.
 DR P-PSDB; AAR91307.

PT Helicobacter pylori Tag A gene - used to develop prods. for the
 PT diagnosis, treatment and prevention of peptic ulceration and gastric
 XX carcinoma.

PS Claim 4; Page 96-102; 118pp; English.

CC AAAT14051 encodes a 120-128 kD antigen of Helicobacter pylori (HP). The
 CC antigen is designated TagA and 1c, or its fragments, can be used to
 CC determine a predilection to peptic ulceration or gastric carcinoma,
 CC conditions caused by or linked to HP infection. TagA and its fragments
 CC may also be used for antibody prodn. for use in detection of TagA in
 CC patients suspected of HP infection. TagA antibodies or other ligands may
 CC also be used to treat peptic ulceration or gastric carcinoma caused by HP

CC infection. TagA or a non-functional TagA mutant may be used in vaccines
 CC for preventing and treating HP infection

QY Sequence 4821 BP; 1721 A; 875 C; 929 G; 1296 T; 0 U; 0 Other;

Query Match 70.8%; Score 485.2; DB 2; Length 4821;

Best Local Similarity 86.6%; Pred. No. 1.3e-109;

Matches 561; Conservative 0; Mismatches 78; Indels 9; Gaps 2;

QY 1 AAAAAATGCAAAATTAAGGATTTCAAGCAAGGAGTACCAAGCAAAAGCACTTGAAT 60
 Db 3319 AAAAAATGCAAAATTAAGGATTTCAAGCAAGGAGTACCAAGCAAAAGCACTTGAAT 3378
 QY 61 TCCGTTAAAGATGTGATCATCAATCAAAAGTAAACGATTAATCTCAATCAA 120
 Db 3379 TCCATTTAAAGATGTGATCATCAATCAAAAGTAAACGATTAATCTCAATCAA 3438
 QY 121 GCGGTATCAGTGTCTTAAGCAAGGCTGTTCAAGTGTGAGCAAGGCTTACCGAT 180
 Db 3439 GCGGTATCAGTGTCTTAAGCAAGGCTGTTCAAGTGTGAGCAAGGCTTACCGAT 3498
 QY 181 CTCAAAAATTTCTCAAAAGCAAGGCTTACCGATCAAGGCTTCAAAATGAAGTCTCAAT 240
 Db 3499 CTCAAAAATTTCTCAAAAGCAAGGCTTACCGATCAAGGCTTCAAAATGAAGTCTCAAT 3558
 QY 241 GCTGAAAAAATTTCTGAAATTTATTCATTCGTTAAGATGTGTGATGAAACCTTATGTC 300
 Db 3559 GTTGG---AAATCTGAAATTTATTCATTCGTTAAGATGTGTGATGAAACCTTATGTC 3615
 QY 301 GGTAAATGTTATCTCAAGCAAGGCTTCTTCTTCAAAATCTTTCGACATCAAG 360
 Db 3616 GGTAAATGTTATCTCAAGCAAGGCTTCTTCTTCAAAATCTTTCGACATCAAG 3675
 QY 361 AAGAGTTGAATGCAAACTTGGAATTTCAATTAACATTAATGAGTCAAAAC 420
 Db 3676 AAGAGTTGAATGCAAACTTGGAATTTCAATTAACATTAATGAGTCAAAAC 3735
 QY 421 -----GAACCATTTATGCTTAAGTTAAATTAAGGAGGCAAGGAGTACCTT 474
 Db 3736 GCGGAGAACCATTTATGCTTAAGTTAAATTAAGGAGGCAAGGAGTACCTT 3795
 QY 475 GAAGAACCATTTATGCTTAAGTTAAATTAAGGAGGCAAGGAGTACCTT 534
 Db 3796 GAAGAACCATTTATGCTTAAGTTAAATTAAGGAGGCAAGGAGTACCTT 3855
 QY 535 CAATTAAGCAATGTTGGGTGTTGTAAGGCAAGGAGGCAAGGAGTACCTT 594
 Db 3856 CAAGCAAGCAAGTGTGTTGTTGTAAGGCAAGGAGGCAAGGAGTACCTT 3915
 QY 595 GATTAAGTTGATGCTCAAGTAAGGAGGCTTCAAGGAGTACCAAGA 642
 Db 3916 GATTAAGTTGATGCTCAAGTAAGGAGGCTTCAAGGAGTACCAAGA 3963

RESULT 10

AAA90022 standard; DNA; 3546 BP.

AAA90022;

DT 20-DEC-2000 (first entry)

XX Cytotoxin associated gene A (caga).

XX Cytotoxin associated gene A; caga; Helicobacter pylori; ds;
 XX digestive tract disease; diagnosis; treatment.

XX Helicobacter pylori.

XX CN1255545-A.

XX 07-JUN-2000.

PF 30-NOV-1998; 98CN-00122056.
 XX
 PR 30-NOV-1998; 98CN-00122056.
 XX
 XX
 PA (JING-) JINGYING BIOTECHNICAL CO LTD SHANGHAI.
 XX
 PI Guo C, Li J, Guo X;
 XX
 DR WPI; 2000-524985/48.
 XX
 PT Process for preparing protein antigen for genetic expression of gene A
 related to *pylorospiribacillus* cytotoxin.
 XX
 PS Disclosure; Fig 1; 16pp; Chinese.
 CC This invention relates to a process for expressing purified *Helicobacter*
 CC *pylori* cytotoxin associated gene A (caga) antigen. The process involves
 CC the use of a plasmid vector to create recombinant caga DNA. The caga
 CC protein can be used to prevent, cure and diagnose diseases of the
 CC digestive tract. The present sequence represents the *Helicobacter pylori*
 CC caga gene
 CC
 SQ Sequence 3546 BP; 1309 A; 627 C; 726 G; 884 T; 0 U; 0 Other;
 Query Match 70.4%; Score 482; DB 3; Length 3546;
 Best Local Similarity 86.3%; Pred. No. 7.2e-109;
 Matches 559; Conservative 0; Mismatches 80; Indels 9; Gaps 2;

QY 1 AAAAAATGGCAAAATTAAGATTTCAGCAAGTACGCAAGCAAAAGCGACCTTGAAT 60
 DB AAAAAATGGCAAAATTAAGATTTCAGCAAGTACGCAAGCAAAAGCGACCTTGAAT 2307
 QY 61 TCCGTTAAGATGATCATCATCAATCAAAAGGTAAGTGAATATTCATCAATCA 120
 DB TCCATTAAGATGATCATCATCAATCAAAAGGTAAGTGAATATTCATCAATCA 2367
 QY 121 GCGGTATCAGTGGCTAAAGCAACGGGTATTCAGTAAAGGTGAGCAAGCGTTAGCCGAT 180
 DB GCGGTATCAGTGGCTAAAGCAACGGGTATTCAGTAAAGGTGAGCAAGCGTTAGCCGAT 2427
 QY 181 CTCAAAAATTTCTCAAGAGCAATGGCCCAACAGCTCAAAAAATGAAAGTCTCAT 240
 DB CTCAAAAATTTCTCAAGAGCAATGGCCCAACAGCTCAAAAAATGAAAGTCTCAT 2487
 QY 241 GCTGAAGAAAAATCTGAATATATCAATCCGTTAAGATGTTGTAATGAAACCTTACTC 300
 DB GCTGAAGAAAAATCTGAATATATCAATCCGTTAAGATGTTGTAATGAAACCTTACTC 2544
 QY 2488 GTTGG--AAATCTGAAATATACCAATCCGTTAAGATGTTGTAAGCAAGCAACCTTACTC 2544
 DB GGTAAATGGATTATCTCAAGAGCAACCACTCTTTCTAAAAATCTTTGGGACATCAAG 360
 QY 2545 GGTAAATGGATTATCTGAATATGAGCCCAAGCTCTCGCCCAAAATTTTCGATATCAAG 2604
 DB 361 AAAGAGTTGAATGCAAACTTGAAATTTCAATTAACAATATATGACTCAAAAC 420
 QY 2605 AAAGAGTTGAATGCAAACTTTAAATTTCAATTAACAATATATGACTCAAAAC 2664
 DB 421 -----GAACCATTTATGCTTAATATTAATAAAGCAAGCGCAAGCTAGCCCTT 474
 QY 2665 GCGGAGAACCATTTATGCTCAAGTTATTAATAAAGCAAGCGCAAGCTAGCCCTT 2724
 DB 475 GAAGAACCATTTATGCTCAAGTTCTAAAGGTAAGTAAATGACCGCACTCAAT 534
 QY 2725 GAAGAACCATTTATGCTCAAGTTCTAAAGGTAAGTAAATGACCGCACTCAAT 2784
 QY 535 CAAATAGCAAGTGGTTGGTGTGTAGGCAAGCAAGCGGCTTCCCTTGAAGAGGAT 594
 DB 2785 CAAAGCAAGCAAGTGGTGTGTGTAGGCAAGCAAGCGGCTTCCCTTGAAGAGGAT 2844
 QY 595 GATAAAGTGAATGATCTCAAGTAAAGTAAAGGCTTTCAAGAAATCAAGAA 642
 DB 2845 GATAAAGTGAATGATCTCAAGTAAAGTAAAGGCTTTCAAGAAATCAAGAA 2892

RESULT 11
 AEB91795
 ID AEB91795 standard; DNA; 3504 BP.
 XX
 AC AEB91795;
 XX
 XX 20-OCT-2005 (first entry)
 DT
 XX
 DE DNA encoding microbial pathogen adhesin protein, SEQ ID NO:504.
 XX
 KW algorithm; adhesin; pharmaceutical; vaccine; drug screening;
 KW *Bordetella pertussis* infection; antibacterial; pneumonia;
 KW anti-inflammatory; respiratory-gen.; gastric ulcer; antitumor;
 KW gastrointestinal-gen.; urinary tract infection; antimicrobial; uropathic;
 KW ds; gene.
 XX
 OS
 XX
 PM WO2005076010-A2.
 PN
 XX
 PD 18-AUG-2005.
 XX
 PF 07-FEB-2005; 2005WO-IN000037.
 XX
 PR 06-FEB-2004; 2004IN-DE000173.
 PR 20-JUL-2004; 2004US-0589227P.
 XX
 PA (COUL) COUNCIL SCI & IND RES SOUTH AFRICA.
 XX
 PI Sachdeva G, Kumar K, Jain P, Brahmachari SK, Ramachandran S;
 DR WPI; 2005-597835/61.
 XX
 PT Computational method for identifying adhesin and adhesin like molecules,
 PT comprises computing sequence-based attributes of protein sequences using
 PT neural network software and training an artificial neural network.
 PS Claim 14; SEQ ID NO 505; 402pp; English.

The present invention relates to a computational method (M1) for
 identifying adhesin and adhesin-like proteins, by computing the sequence-
 based attributes of protein sequences using five attribute modules of a
 neural network software, training an artificial neural network (ANN) for
 each of the computed five attributes, and identifying the adhesin and
 adhesin-like proteins having probability of being an adhesin (pad) as
 equal or greater than 0.51. Also claimed is a set of 274 annotated genes
 encoding adhesin and adhesin-like proteins, having 274 fully defined 162-
 1151 base pair (SEQ ID NO: 385-658) sequences; a set of 105 hypothetical
 genes encoding adhesin and adhesin-like proteins, having 105 fully
 defined 306-15876 base pairs (SEQ ID NO: 659-763); a set of 279 annotated
 adhesin and adhesin-like proteins, having 279 fully defined 53-3716 base
 pair (SEQ ID NO: 1-279); a set of 105 hypothetical adhesin and adhesin-
 like proteins, having 105 fully defined 106-5291 base pair (SEQ ID NO:
 280-384) sequences; and a fully connected multilayer feed forward ANN (I)
 based on (M1). (M1) is useful for identifying adhesin and adhesin-like
 proteins, of therapeutic potential, and identifying and short-listing
 proteins for further testing in development of new vaccine formulations
 to eliminate diseases caused by various pathogenic organisms. (M1) is
 useful for identifying putative adhesins that are important in drug
 discovery and preventing therapeutics for whooping cough, pneumonia,
 gastric ulcer and urinary tract infections. (M1) identifies adhesins from
 CC distantly related organisms, and from bacteria belonging to a wide
 CC phylogenetic spectrum. (M1) is capable of predicting adhesive nature of
 CC unique proteins. The present sequence is DNA encoding a microbial
 CC pathogen adhesin protein sequence.
 CC
 SQ Sequence 3504 BP; 1293 A; 600 C; 715 G; 896 T; 0 U; 0 Other;
 Query Match 69.7%; Score 477.2; DB 14; Length 3504;
 Best Local Similarity 86.9%; Pred. No. 1.1e-107;
 Matches 558; Conservative 0; Mismatches 33; Indels 51; Gaps 1;

QY 1 AAAAAATGGCAAAATTAAGATTTCAGCAAGTACGCAAGCAAAAGCGACCTTGAAT 60

	RESULT	12
	AAT67410	
ID	AAT67410	standard; DNA, 3534 BP.
XX		
AC	AAT67410,	
XX		
DT	08-JUL-1997	(first entry)
DE		
XX	H. pylori cytoplasmic protein ORF 16459375..aa.	
XX	Cytoplasmic; vaccine, prevention; infection; identification;	
KW	binding compound; bacterium, life cycle; activator; bacteria; inhibitor;	
KX	duodenal ulcer disease; chronic gastritis; diagnosis; envelope, ds.	
XX		
OS	Helicobacter pylori.	
XX		
PFH	Key	Location/Qualifiers
PFT	CDS	1..3534
PFT		/*tag= a
PFT		/note= "no stop codon is given"
XXX		
RPN	M09640893-A1.	
XX		
PD	19-DEC-1996.	
XX		
PP	06-JUN-1996;	96MO-US009122.
XX		
PR	07-JUN-1995;	95US-00487032.
OR	01-APR-1996;	96US-00630405.

The present sequence encodes a Helicobacter pylori cytoplasmic protein. The protein may be used in a vaccine to prevent or treat *H. pylori* infection or to identify *H. pylori* polypeptide binding compounds, useful as potential *H. pylori* life cycle activators or inhibitors. The genomic sequence of *H. pylori* (ATCC 55679) was determined from overlapping sequences generated by mechanically shearing the bacterial DNA. The contigs were analysed for ORF of at least 180 nucleotides, and the predicted coding regions defined by computer evaluation. To identify likely *H. pylori* antigens for vaccine development, the amino acid sequences predicted from various ORF were analysed for significant homology to other known or exported membrane proteins. Having identified and determined the sequences of interest, particular regions can be isolated from *H. pylori* by PCR amplification for recombinant polypeptide production, e.g. in *E. coli* hosts

Query Match	Best Local Similarity	65.7%	Score 477.2;	DB 2;	Length 3534;
Matches 558;	Conservative	9.9%	Prod. No. 1.1e-107;	Mismatches 307;	Indels 51; Gaps 1
Qy	1	AAAAATGGCAAAATTAAGCATTTGACGAGGTATCCGACAGCAAAAGCCACTTGGAAT	60		
Db	2293	AAAATGGCAAAATTAAGCATTTGACGAGGTATCCGACAGCAAAAGCCACTTGGAAT	2352		
Qy	61	TCCGTTAAAGATGTATCATCATCAATCAAAAGGTAAAGGTATATTCATCA	120		
Db	2353	TCCATTTAAAGATGTATCATCATCAAAAGGTAAAGGTATATTCATCA	2412		
Qy	121	GGGTATCAAGTGGCTAAAGCAAGGGTATTTGATGAGGTAAAGCAACGTTAGCGAT	180		
Db	2413	GCGGTATCAAGTGGCTAAAGCAAGGGTATTTGATGAGGTAAAGCAACGTTAGCGAT	2472		
Qy	181	CTCAAAATTTCTCAAAAGCAATTTGGCTCAACAAGCTCAAAAAATGAAGTCCAT	240		
Db	2473	CTCAAAATTTCTCAAAAGCAATTTGGCTCAACAAGCTCAAAAAATGAAGTTCAT	2532		
Qy	241	GCTGAAGAAAAATCTGAATATATTCATCCGTTAAGATGSTRGAATGGAACCTATGTC	300		
Db	2533	AGTGAAGAAAAATCTGACATATCCATCCGTTAAGATGSTRGAATGGAACCTATGTC	2592		
Qy	301	GGTATGSGTTATCTCAAGCAAGACCAACCTCTTTCTAAAACTTTTGGACATCAAG	360		
Db	2593	GGTATGSGTTATCTCAAGCAAGACCAACCTCTTTCTAAAACTTTTGGACATCAAG	2652		
Qy	361	AAAGATGTAATGCMAAACCTTGGAAATTTCAATTAACAATTAATGAATCAAAAAC	420		
Db	2653	AAAGATGTAATGCMAAACCTTGGAAATTTCAATTAACAATTAATGAATCAAAAAC	2712		
Qy	421	GAACCATTTATGCTAAAGTTAATAAAAAGAAACAGGCGAAGCAAGCTTAGCTTGAAAG	480		
Db	2713	AGCAC-----AATA-2721			
Qy	481	CCCATTTACGCTCAAGTTGCTAAAGAGTAAATGCAAAAATGACCACTCAATCAATA	540		
Db	2722	CCCATTTATCAAGTTGCTAAAGAGTAAAGAGTAAATGACCACTCAATCAATA	2781		
Qy	541	GCATGTGTTGGGTGTGTAAGGCAGACAGGCGCTTCCTTTGAAAGAGCATATTA	600		

QY 1 AAAAATGCCAAAATAAGGATTCAGCAAGGTAACGCAAGCAAAAAGCGACCTTGAAAT 60

DR WPI; 2002-674910/72.
DB R-PSDB: ABUE1384

DR P-PSDB; ABU51384.

XX New complexes of protein-protein interactions in *Helicobacter pylori*,
PT useful for identifying modulating compounds for treating or preventing
XX ulcers in mammals.
PS Claim 7, Page 262; 642pp; English.
CC The invention describes a complex of protein-protein interactions in
CC *Helicobacter pylori* selected from 421 complexes given in the
CC specification. The complex of protein-protein interactions are useful for
CC screening for agents which modulate the interaction of proteins.
CC Modulating compounds which bind to a targeted bacterial protein may be
CC used for treating or preventing ulcers in a human or animal. This
CC sequence encodes a selected interacting domain (SID), identified via
XX protein-protein interactions
SQ Sequence 1024 BP; 381 A; 177 C; 214 G; 252 T; 0 U; 0 Other;
Query Match 59.3%; Score 406.2; DB 6; Length 1024;
Best Local Similarity 87.8%; Pred. No. 2,76-90;
Matches 455; Conservative 0; Mismatches 60; Indels 3; Gaps 1;
QY 128 CAGTGGCTAAGCAAGCGGTATTCAGTGGTGAAGCAAGCGTTAGCCATCTCAAAA 187
DB 1 CAGTGGCTAAGCAAGCGGTATTCAGTGGTGAAGCAAGCGTTAGCCATCTCAAAA 60
QY 188 ATTTCTCAAGAGAGCAATTTGGCCCAAGCTCAAAAAAATGAAAGTCTCAATCTGAA 247
DB 61 ACTTCTCAAGAGAGCAATTTGGCTCAACAGCTCAAAAAAATGAAATTTCAATCTGAA 120
QY 248 AAAATCTGAATATATATCAATCGTTAAGAGGTGGAATGGAACCTTACGTGGTATG 307
DB 121 AAAATCTGAATATATCAATCGTTAAGAGGTGGAATGGAACCTTACGTGGTATG 180
QY 308 GGTATCTCAAGCAAGAGCCCAACTCTTCTAATAAATCTTGGACATCAAGAAAGT 367
DB 181 GGTATCTGAATATAGAGCCCAAGCTCTCGCAAAATTTTGGATATCAAGAAAGT 240
QY 368 TGAATGCAAACTTGGAAATTTTCAATTAACATTAACAT--AATGATCTCAAAAGCA 424
DB 241 TGAATGCAAAATTTTAAATTTTCAATTAACATTAATGAGCTCAAAAGCAAGCA 300
QY 425 CCATTAATGCTAAGTATTAATAAAGAAAGAGGAGGAGCAAGCTGAGCAAGCA 484
DB 301 CCATTAATGCTAAGTATTAATAAAGAAAGAAAGAGCAAGCAAGTACCTGAGCA 360
QY 485 TTTAGCTCAAGTGTCTAATAAAGTAAATGCAAAATTAAGCAAGCTCAATCAATGCA 544
DB 361 TTTAGCTCAAGTGTCTAATAAAGTAAATGCAAAATTAAGCAAGCTCAATCAATGCA 420
QY 545 GTGGTTGGGTGTGTATGAGGCAAGCAAGCGGCTTCCCTTGAAGAGCATGAATG 604
DB 421 GTGGTTGGGTGTGTATGAGGCAAGCAAGCGGCTTCCCTTGAAGAGCATGAATG 604
QY 605 ATGATCTCAAGTAAAGTAAAGGCTTTTCAAGCAATCAAGAA 642
DB 481 ATGATCTCAAGTAAAGTAAAGGCTTTTCAAGTAAAGCTTGA 518
RESULT 15
ABX66698 standard; DNA; 854 BP.
XX ABX66698;
AC
XX
XX 07-MAY-2003 (first entry)
DB *Helicobacter pylori* selected interacting domain (SID) DNA #1297.
XX
XX Protein-protein interaction; ulcer; selected interacting domain; SID;
KM gene; db.
XX
OS *Helicobacter pylori*.

XX WO200266501-A2.
XX
XX 29-AUG-2002.
XX
XX 28-DEC-2001; 2001WO-EP015428.
XX
XX 02-JAN-2001; 2001US-0259302P.
XX
XX (HYBR-) HYBRISNETS.
XX (INSP) INST PASTEUR.
XX
XX Legrain P, Rain J, Collard F, De Reuse H, Labigne A;
XX WPI; 2002-674910/72.
XX
XX P-PSDB; ABUS1954.
XX
XX New complexes of protein-protein interactions in *Helicobacter pylori*,
PT useful for identifying modulating compounds for treating or preventing
XX ulcers in mammals.
PS Claim 7, Page 399; 642pp; English.
CC The invention describes a complex of protein-protein interactions in
CC *Helicobacter pylori* selected from 421 complexes given in the
CC specification. The complex of protein-protein interactions are useful for
CC screening for agents which modulate the interaction of proteins.
CC Modulating compounds which bind to a targeted bacterial protein may be
CC used for treating or preventing ulcers in a human or animal. This
CC sequence encodes a selected interacting domain (SID), identified via
XX protein-protein interactions
SQ Sequence 854 BP; 328 A; 150 C; 174 G; 202 T; 0 U; 0 Other;
Query Match 55.4%; Score 379.4; DB 6; Length 854;
Best Local Similarity 87.5%; Pred. No. 1,16-85;
Matches 427; Conservative 0; Mismatches 58; Indels 3; Gaps 1;
QY 158 GGGTAAAGCAAGCGGTATTCAGTGGTGAAGCAAGCGTTAGCCATCTCAAAA 217
DB 1 GGGTAAAGCAAGCGGTATTCAGTGGTGAAGCAAGCGTTAGCCATCTCAAAA 60
QY 218 CTCAAAAAATGAAGTCTCAATGCTAAGAAAAATCTGAATATATCAATCTGTTAAGA 277
DB 61 CTCAAAAAATGAAGTCTCAATGCTAAGAAAAATCTGAATATATCAATCTGTTAAGA 120
QY 278 ATGATGAAATGAAGCCCTAGTGGTAAATGGAATGCAAGCAAGCAAGCTCTT 337
DB 121 ATGATGAAATGAAGCCCTAGTGGTAAATGGAATGCAAGCAAGCTCTT 180
QY 338 CTAAAAAATCTTTCGCAATCAAGAAAGGTTGAATGCAAAATCTGAAATTTCAATCA 397
DB 181 CAAAAAATTTTTCGCAATCAAGAAAGGTTGAATGCAAAATTTCAATCA 240
QY 398 ATTAACAT--AATGACTCAAAAAGCAAGCCATTAAGCTAAGTAAATTAAGAAAG 454
DB 241 ATTAATTAAGCTCAAAAAGCAAGCCATTAAGCTAAGTAAATTAAGAAAG 300
QY 455 CAGGCAAGCAAGCTAGCTTGAAGCAAGCCATTAAGCTAAGTAAATTAAGAAAG 514
DB 301 CAGGCAAGCTAGCTAGCTTGAAGCAAGCCATTAAGCTAAGTAAATTAAGAAAG 360
QY 515 CAAAAATTAAGCAAGCTCAATCAATGAAGTAAAGTAAAGTAAAGTAAAGTAAAG 574
DB 361 CAAAAATTAAGCAAGCTCAATCAATGAAGTAAAGTAAAGTAAAGTAAAGTAAAG 420
QY 575 GCTTCCCTTGAAGAGCATGAATTAAGTAAATCTAGTAAGTAAAGTAAAGTAAAG 634
DB 421 GCTTCCCTTGAAGAGCATGAATTAAGTAAATCTAGTAAGTAAAGTAAAGTAAAG 634
QY 635 ATCAAGAA 642
DB 481 GCCCTGAA 488

Tue Mar 7 12:58:25 2006

Search completed: March 6, 2006, 22:05:53
Job time : 558 secs

us-09-360-685c-26.rng

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2006, 21:27:13 ; Search time 3660 Seconds

(without alignments)
10638.733 Million cell updates/sec

Title: US-09-360-685C-26

Perfect score: 685
Sequence: 1 aaatggcaataaagga.....caatcaagcgatcagaag 685

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_in: *
3: gb_env: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pr: *
9: gb_ro: *
10: gb_ste: *
11: gb_sy: *
12: gb_un: *
13: gb_vl: *
14: gb_hyg: *
15: gb_pl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	685	100.0	5925	1 HPCAI	X70039 H. pylori ca
2	685	100.0	5925	6 A76409	A76409 Sequence 4
3	685	100.0	5925	6 AR099343	AR099343 Sequence
4	685	100.0	5925	6 AR112380	AR112380 Sequence
5	685	100.0	5925	6 AX030324	AX030324 Sequence
6	685	100.0	5925	6 BD000835	BD000835 Helicobac
7	685	100.0	5925	6 BD000839	BD000839 Helicobac
8	685	100.0	5925	6 AF282853	AF282853 Helicobac
9	685	100.0	5925	6 AB015416	AB015416 Helicobac
10	601.8	87.9	3426	1 DQ011619	DQ011619 Helicobac
11	597	87.2	978	1 AB110963	AB110963 Helicobac
12	595.6	86.9	3693	1 DQ067454	DQ067454 Helicobac
13	589.2	86.0	26663	1 AY33063852	AY33063852 Helicobac
14	584.4	85.3	3543	1 AB090088	AB090088 Helicobac
15	584.4	85.3	26771	1 AY330637	AY330637 Helicobac
16	584.4	85.3	37037	1 AB120425	AB120425 Helicobac
17	579.4	84.6	981	1 AB090151	AB090151 Helicobac
18	579.4	84.6	3447	1 AB190949	AB190949 Helicobac

19	570	83.2	3540	1 AB015414	AB015414 Helicobac
20	566.6	82.7	815	6 AR679176	AR679176 Sequence
21	565.2	82.5	3540	1 AF247651	AF247651 Helicobac
22	565	82.5	801	1 AF479032	AF479032 Helicobac
23	558.4	81.5	3471	1 AY884088	AY884088 Helicobac
24	554	80.9	3744	1 AF202973	AF202973 Helicobac
25	552.4	80.6	3744	1 AB015413	AB015413 Helicobac
26	549.2	80.2	3744	1 AB003387	AB003387 Helicobac
27	549.2	80.2	3744	1 AF001357	AF001357 Helicobac
28	547.6	79.9	7498	1 AY33064053	AY33064053 Helicobac
29	547.6	79.9	26276	1 AF23064352	AF23064352 Helicobac
30	544.2	79.4	1323	1 AF222809	AF222809 Helicobac
31	541.2	79.0	4093	1 AF083352	AF083352 Helicobac
32	539.6	78.8	1077	1 AB090143	AB090143 Helicobac
33	539.6	78.8	3552	1 AB190940	AB190940 Helicobac
34	534.8	78.1	933	1 AY571191	AY571191 Helicobac
35	534.8	78.1	933	1 AY571192	AY571192 Helicobac
36	534.8	78.1	933	1 AY571193	AY571193 Helicobac
37	534.8	78.1	933	1 AY571194	AY571194 Helicobac
38	534.8	78.1	933	1 AY571195	AY571195 Helicobac
39	534.8	78.1	933	1 AY571196	AY571196 Helicobac
40	534.4	78.0	1224	1 AF222808	AF222808 Helicobac
41	532.4	77.7	3558	1 DQ011620	DQ011620 Helicobac
42	531.6	77.6	3558	6 CQ903292	CQ903292 Sequence
43	531.6	77.6	3551	1 AB090082	AB090082 Helicobac
44	531.6	77.6	11141	1 AB000569	AB000569 Helicobac
45	531.6	77.6	37459	1 AB120420	AB120420 Helicobac

ALIGNMENTS

RESULT 1	HPCAI	5925 bp	DNA	linear	BCT 22-JUL-1993
LOCUS	H. pylori cai gene for cytotoxicity associated immunodominant antigen.				
DEFINITION	H. pylori cai gene for cytotoxicity associated immunodominant antigen.				
ACCESSION	X70039.1	GI:394912			
VERSION	X70039				
KEYWORDS	cytotoxicity associated immunodominant antigen.				
SOURCE	Helicobacter pylori				
ORGANISM	Helicobacter pylori				
REFERENCE	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales; Helicobacteraceae; Helicobacter.				
AUTHORS	Covacci, A., Censini, S., Bugnoli, M., Petracca, R., Burroni, D., Macchia, G., Massone, A., Papini, E., Xiang, Z., Figura, N. and Rapunoli, R.				
TITLE	Molecular characterization of the 128-kDa immunodominant antigen of Helicobacter pylori associated with cytotoxicity and duodenal ulcer				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 90 (12), 5791-5795 (1993)				
PUBMED	8516329				
REFERENCE	2 (bases 1 to 5925)				
AUTHORS	Covacci, A.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-JAN-1993) A. Covacci, Irtis, Immunobiological Research Inst Siena, Via Fiorentina, 1, 53100 Siena, ITALY				
FEATURES	Location/Qualifiers				
source	1..5925				
	/organism="Helicobacter pylori"				
	/mol_type="genomic DNA"				
	/strain="CCUG 17874"				
	/db_xref="taxon:210"				
	/clone="57/D.64/4. 007, 24, A1, G5, A17, B1"				
	/clone_lib="lambda gtl1 AR7/plasmid pINDIII"				
	1..534				
	promoter				
	prim_transcript 324..4032				
	/note="transcript B"				
	430..4032				
	/note="transcript A"				
	prim_transcript 430..4032				
	/note="transcript A"				
	430..4032				
	RBS				
	525..529				
	535..3878				
	gene				
	/gene="cai"				

CDS

```
535..3978
/ gene="cal"
/ codon_start=1
/ translatable=1
/ product="cytotoxicity associated immunodominant antigen"
/ protein_id="CAA49633.1"
/ db_xref="GI:394913"
/ db_xref="GOA:880200"
/ db_xref="UniProt/Swiss-Prot:P80200"
/ translation="MTNETIDQPPQEAFAFPQPIINILQVAFKLVNVAASYPDPK
PIVDKNDNRNRAEFGISQIREVYSKAIKNPTEKQNFQDFINKNLIDV
ESSTKSFQFQFQRIPTSMVHONDPKINTSINEMENIOPPLIDKKRQAPL
KSAKOSFAGIIGNOIRTDQFQVDESLERQEAKEGEPGGDMIDPLSFIDK
KQSDVKEALINDEVPVQPDIAITTTDIOGLPEARDLLDRGNFSFTLGMEMLD
VEGVADIDPNYKFNQLIHNNALSVLMSHNGIPEKSLLYGGNGGGRARRDMAT
VGYKDOQNNVATTIINVMKNGSGVIAAGKGINNPSFLYKEDQLGSGRALSOE
IQNKIDPMEFLQNNALDNLSEKEKEFTEIKDFQKSKAYLDALGNRIAPVSK
DTGSHALITEFGNDLSYLLKQYKADKALDRKQNTIQQSLKHDVMEVDYSNKY
TNASKPNKGVGVNGVSHLEVGNKVAIPFLPDINNLITSPRRNLEKLTREGS
PEEANKLIDPLSSNKELVGKTINPKVAIDAKVQNGVDYKQKQDLFEKIKRHL
EKEVEKLESKGNKMEKAQANSODEIPALINKENPRAPALYVONLGIKRE
LSDKLENVQKLDPKSPDEPNKNDPSKAEETLKLSQYDQGINPEKIRE
NLNAILNPKNGKDKSKVTOAKSDLENSVQVITINQYPTKDNQNVAVAKRG
DFERVQALADLKNSKEQALQAKNSODEIKYQSVKNGVNTQVAGVSAQ
EATLSKPNSDIKKELNAKGNPNNNNNGIKNEPIYQSVKNGVNTQVAGVSAQ
VAKNNAKIDLNQIASGIVGACAPPLRDKTDVLSKGLSKQELQALIDNLA
QAVSEKAGFPNGLEOTIDKLDSTKHPMLVSEAKVPSASLADLVANSHIR
INSNTKNAINERKATMLTKRPPKLVNDKIVARHVSFLSYDKLGFQNKMKD
YSDSPFSTKNNAVNDTNSGTFPLTNAISTASYICLARENABEHKVNTRKGFQK
S"
terminator
4010..4032
```

Query Match 100.0%; Score 685; DB 1; Length 5925;
Best Local Similarity 100.0%; Pred. No. 5.2e-118;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AAAAATGGCAAAATTAAGATTTCAGCAAGTAAACGCAAGCAAAAGCACTTGAAT 60
Db 2782 AAAAATGGCAAAATTAAGATTTCAGCAAGTAAACGCAAGCAAAAGCACTTGAAT 2841
QY 61 TCCGTTAAAGATGATCATCAATCAAAAGTAAACGATTAAGTATATCTCAATCA 120
Db 2842 TCCGTTAAAGATGATCATCAATCAAAAGTAAACGATTAAGTATATCTCAATCA 2901
QY 121 GCGGTATCAGTGGCTAAAGCAAGGATGATTCAGTAGGCTAGAGCAAGCTTACCGAT 180
Db 2902 GCGGTATCAGTGGCTAAAGCAAGGATGATTCAGTAGGCTAGAGCAAGCTTACCGAT 2961
QY 181 CTCAAAAATTTCTCAAGAGCAATGGCCCAACAGCTCAAAAAATTAAGAACTTCAT 240
Db 2962 CTCAAAAATTTCTCAAGAGCAATGGCCCAACAGCTCAAAAAATTAAGAACTTCAT 3021
QY 241 GCTAGAAAAAATCTGAATAATATATCAATCCGTTAAGATGATGTAATGAACCTTAC 300
Db 3022 GCTAGAAAAAATCTGAATAATATATCAATCCGTTAAGATGATGTAATGAACCTTAC 3081
QY 301 GGTAAATGGGTATCTCAAGAGCAAGCAACCTTTCTTAAAACTTTTGGACATCAAG 360
Db 3082 GGTAAATGGGTATCTCAAGAGCAAGCAACCTTTCTTAAAACTTTTGGACATCAAG 3141
QY 361 AAAGAGTTGAATGCAAACTTGAATTTCAATACATTAATGAATGCACTCAAAAAC 420
Db 3142 AAAGAGTTGAATGCAAACTTGAATTTCAATACATTAATGAATGCACTCAAAAAC 3201
QY 421 GAACCATTTATGCTAAAGTTAATAAAGAAAGCAAGGCAAGCAAGCTTACCTTGAAGA 480
Db 3202 GAACCATTTATGCTAAAGTTAATAAAGAAAGCAAGGCAAGCAAGCTTACCTTGAAGA 3261
QY 481 CCATTTAGCTCAAGTTGCTAAAGGTAATGCAAAATTTGACCGACTCAATCAATA 540
Db 3262 CCATTTAGCTCAAGTTGCTAAAGGTAATGCAAAATTTGACCGACTCAATCAATA 3321
```

```
QY 541 GCAAGTGTGTTGGTGTGTTGTAAGGCAAGCGGCTTCCCTTTGAAAAGCAATGATA 600
Db 3322 GCAAGTGTGTTGGTGTGTTGTAAGGCAAGCGGCTTCCCTTTGAAAAGCAATGATA 3381
QY 601 GTTGAATGATCTCAATTAAGTAAAGCTTTCAAGATATCAAGAAATGGCTCAGAAAATGAC 660
Db 3382 GTTGAATGATCTCAATTAAGTAAAGCTTTCAAGATATCAAGAAATGGCTCAGAAAATGAC 3441
QY 661 AATTCATTCAGCGGCTATCAGAG 685
Db 3442 AATTCATTCAGCGGCTATCAGAG 3466
```

RESULT 2
LOCUS A76409
DEFINITION Sequence 4 from Patent WO9318150.
ACCESSION A76409
VERSION A76409.1 GI:6088449
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 5925)
Covacci, A. and Bugnoli, M.
TITLE
HELI-CO-BACTER PYLORI PROTEINS USEFUL FOR VACCINES AND DIAGNOSTICS
JOURNAL
PATENT: WO 9318150-A 4 16-SEP-1993;
SCLAVO BIOCHIM SPA (IT)
FEATURES
source
1..5925
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"

ORIGIN

Query Match 100.0%; Score 685; DB 6; Length 5925;
Best Local Similarity 100.0%; Pred. No. 5.2e-118;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AAAAATGGCAAAATTAAGATTTCAGCAAGTAAACGCAAGCAAAAGCACTTGAAT 60
Db 2782 AAAAATGGCAAAATTAAGATTTCAGCAAGTAAACGCAAGCAAAAGCACTTGAAT 2841
QY 61 TCCGTTAAAGATGATCATCAATCAAAAGTAAACGATTAAGTATATCTCAATCA 120
Db 2842 TCCGTTAAAGATGATCATCAATCAAAAGTAAACGATTAAGTATATCTCAATCA 2901
QY 121 GCGGTATCAGTGGCTAAAGCAAGGATGATTCAGTAGGCTAGAGCAAGCTTACCGAT 180
Db 2902 GCGGTATCAGTGGCTAAAGCAAGGATGATTCAGTAGGCTAGAGCAAGCTTACCGAT 2961
QY 181 CTCAAAAATTTCTCAAGAGCAATGGCCCAACAGCTCAAAAAATTAAGAACTTCAT 240
Db 2962 CTCAAAAATTTCTCAAGAGCAATGGCCCAACAGCTCAAAAAATTAAGAACTTCAT 3021
QY 241 GCTAGAAAAAATCTGAATAATATATCAATCCGTTAAGATGATGTAATGAACCTTAC 300
Db 3022 GCTAGAAAAAATCTGAATAATATATCAATCCGTTAAGATGATGTAATGAACCTTAC 3081
QY 301 GGTAAATGGGTATCTCAAGAGCAAGCAACCTTTCTTAAAACTTTTGGACATCAAG 360
Db 3082 GGTAAATGGGTATCTCAAGAGCAAGCAACCTTTCTTAAAACTTTTGGACATCAAG 3141
QY 361 AAAGAGTTGAATGCAAACTTGAATTTCAATACATTAATGAATGCACTCAAAAAC 420
Db 3142 AAAGAGTTGAATGCAAACTTGAATTTCAATACATTAATGAATGCACTCAAAAAC 3201
QY 421 GAACCATTTATGCTAAAGTTAATAAAGAAAGCAAGGCAAGCAAGCTTACCTTGAAGA 480
Db 3202 GAACCATTTATGCTAAAGTTAATAAAGAAAGCAAGGCAAGCAAGCTTACCTTGAAGA 3261
QY 481 CCATTTAGCTCAAGTTGCTAAAGGTAATGCAAAATTTGACCGACTCAATCAATA 540
Db 3262 CCATTTAGCTCAAGTTGCTAAAGGTAATGCAAAATTTGACCGACTCAATCAATA 3321
```

Db 3262 CCATTTACGCTCAAGTTGCTAAAGTAATGCAAAATTTGACCACTCAATCAATA 3321

Qy 541 GCAAGTGTGGGTGTTGTAGAGGCAAGCGGGCTTCCCTTTGAAAAGCATGATATA 600

Db 3322 GCAAGTGTGGGTGTTGTAGAGGCAAGCGGGCTTCCCTTTGAAAAGCATGATATA 3381

Qy 601 GTTGATGATCTCAAGTAAAGGCTTTCAAGAAATCAAGAAATGGCTCAGAAAATGAC 660

Db 3382 GTTGATGATCTCAAGTAAAGGCTTTCAAGAAATCAAGAAATGGCTCAGAAAATGAC 3441

Qy 661 AATCTCAATCAAGCGGTATCAGAG 685

Db 3442 AATCTCAATCAAGCGGTATCAGAG 3466

RESULT 3

AR099343 5925 bp DNA linear PAT 14-FEB-2001

LOCUS AR099343

DEFINITION Sequence 4 from patent US 6077706.

ACCESSION AR099343

VERSION AR099343.1 GI:12809109

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 5925)

AUTHORS Covacci, A., Bugnoli, M., Telford, J., Macchia, G. and Rappuoli, R.

TITLE Helicobacter pylori proteins useful for vaccines and diagnostics

JOURNAL Patent: US 6077706-A 4 20-JUN-2000;

FEATURES

source 1..5925

/organism="unknown"

/mol_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 685; DB 6; Length 5925;

Best Local Similarity 100.0%; Pred. No. 5.2e-118;

Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAGGCAAAATTAAGATTTCAGCAAGTAAAGCAAAAGCAAGCACTTGAAAT 60

Db 2782 AAAAAAGGCAAAATTAAGATTTCAGCAAGTAAAGCAAAAGCAAGCACTTGAAAT 2841

Qy 61 TCCGTTAAAGATGTATCATCAATCAAAAGTAAAGCAATTAATGATATCTCATCA 120

Db 2842 TCCGTTAAAGATGTATCATCAATCAAAAGTAAAGCAATTAATGATATCTCATCA 2901

Qy 121 GCGGTATCAGTGTCTAAAGCAAGGCTGATTTCACTAGGCTAGAGCAAGGCTTAGCCGAT 180

Db 2902 GCGGTATCAGTGTCTAAAGCAAGGCTGATTTCACTAGGCTAGAGCAAGGCTTAGCCGAT 2961

Qy 181 CTCAAAAATTTCTCAAGAGCAATTTGGCCCAAGCTCAAAAAATGAAAGTCTCAAT 240

Db 2962 CTCAAAAATTTCTCAAGAGCAATTTGGCCCAAGCTCAAAAAATGAAAGTCTCAAT 3021

Qy 241 GCTAGAAAAAATCTGAATATATCATCCGTTAAGAAATGCTGATGATGAAACCTTAGTC 300

Db 3022 GCTAGAAAAAATCTGAATATATCATCCGTTAAGAAATGCTGATGATGAAACCTTAGTC 3081

Qy 301 GGTAAATGGTTATCTCAAGCAAGCAAGCAACTCTTTTAAAAATCTTTGCGACATCAAG 360

Db 3082 GGTAAATGGTTATCTCAAGCAAGCAAGCAACTCTTTTAAAAATCTTTGCGACATCAAG 3141

Qy 361 AAGAGTTGAATGCAAACTTGAAATTTCAATTAACATTAACATTAATGATGACTCAAAAC 420

Db 3142 AAGAGTTGAATGCAAACTTGAAATTTCAATTAACATTAACATTAATGATGACTCAAAAC 3201

Qy 421 GAACCATTTATGCTAAAGTATAAAGAAAGCAAGGCAAGGCAAGCTAGCTTGAGAA 480

Db 3202 GAACCATTTATGCTAAAGTATAAAGAAAGCAAGGCAAGGCAAGCTAGCTTGAGAA 3261

Qy 481 CCCATTAGCTCAAGTGTCTAAAGAAAGTAAATGCAAAATTTGACCGACTCAATCAATA 540

Db 3262 CCATTTACGCTCAAGTTGCTAAAGTAATGCAAAATTTGACCACTCAATCAATA 3321

Qy 541 GCAAGTGTGGGTGTTGTAGAGGCAAGCGGGCTTCCCTTTGAAAAGCATGATATA 600

Db 3322 GCAAGTGTGGGTGTTGTAGAGGCAAGCGGGCTTCCCTTTGAAAAGCATGATATA 3381

Qy 601 GTTGATGATCTCAAGTAAAGGCTTTCAAGAAATCAAGAAATGGCTCAGAAAATGAC 660

Db 3382 GTTGATGATCTCAAGTAAAGGCTTTCAAGAAATCAAGAAATGGCTCAGAAAATGAC 3441

Qy 661 AATCTCAATCAAGCGGTATCAGAG 685

Db 3442 AATCTCAATCAAGCGGTATCAGAG 3466

RESULT 4

AR112380 5925 bp DNA linear PAT 16-MAY-2001

LOCUS AR112380

DEFINITION Sequence 4 from patent US 6130059.

ACCESSION AR112380

VERSION AR112380.1 GI:14092280

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 5925)

AUTHORS Covacci, A., Bugnoli, M., Telford, J., Macchia, G. and Rappuoli, R.

TITLE Helicobacter pylori proteins useful for vaccines and diagnostics

JOURNAL Patent: US 6130059-A 4 10-OCT-2000;

FEATURES

source 1..5925

/organism="unknown"

/mol_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 685; DB 6; Length 5925;

Best Local Similarity 100.0%; Pred. No. 5.2e-118;

Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAGGCAAAATTAAGATTTCAGCAAGTAAAGCAAAAGCAAGCACTTGAAAT 60

Db 2782 AAAAAAGGCAAAATTAAGATTTCAGCAAGTAAAGCAAAAGCAAGCACTTGAAAT 2841

Qy 61 TCCGTTAAAGATGTATCATCAATCAAAAGTAAAGCAATTAATGATATCTCATCA 120

Db 2842 TCCGTTAAAGATGTATCATCAATCAAAAGTAAAGCAATTAATGATATCTCATCA 2901

Qy 121 GCGGTATCAGTGTCTAAAGCAAGGCTGATTTCACTAGGCTAGAGCAAGGCTTAGCCGAT 180

Db 2902 GCGGTATCAGTGTCTAAAGCAAGGCTGATTTCACTAGGCTAGAGCAAGGCTTAGCCGAT 2961

Qy 181 CTCAAAAATTTCTCAAGAGCAATTTGGCCCAAGCTCAAAAAATGAAAGTCTCAAT 240

Db 2962 CTCAAAAATTTCTCAAGAGCAATTTGGCCCAAGCTCAAAAAATGAAAGTCTCAAT 3021

Qy 241 GCTAGAAAAAATCTGAATATATCATCCGTTAAGAAATGCTGATGATGAAACCTTAGTC 300

Db 3022 GCTAGAAAAAATCTGAATATATCATCCGTTAAGAAATGCTGATGATGAAACCTTAGTC 3081

Qy 301 GGTAAATGGTTATCTCAAGCAAGCAAGCAACTCTTTTAAAAATCTTTGCGACATCAAG 360

Db 3082 GGTAAATGGTTATCTCAAGCAAGCAAGCAACTCTTTTAAAAATCTTTGCGACATCAAG 3141

Qy 361 AAGAGTTGAATGCAAACTTGAAATTTCAATTAACATTAACATTAATGATGACTCAAAAC 420

Db 3142 AAGAGTTGAATGCAAACTTGAAATTTCAATTAACATTAACATTAATGATGACTCAAAAC 3201

Qy 421 GAACCATTTATGCTAAAGTATAAAGAAAGCAAGGCAAGGCAAGCTAGCTTGAGAA 480

Db 3202 GAACCATTTATGCTAAAGTATAAAGAAAGCAAGGCAAGGCAAGCTAGCTTGAGAA 3261

Qy 481 CCCATTAGCTCAAGTGTCTAAAGAAAGTAAATGCAAAATTTGACCGACTCAATCAATA 540

```

Db      3462  CCATTTAGGCTCAAGTTGCTAAAAAGTTAAATGCAAAAATTGACCACTCAATCAATA 3321
Qy      541  GCAAGTGGTTGGTGTGTGTAAGGGCAAGCAGCGGGCTTCCCTTGAAGAAAGCATGATAA 600
Db      3322  GCAAGTGGTTGGTGTGTGTAAGGGCAAGCAGCGGGCTTCCCTTGAAGAAAGCATGATAA 3381
Qy      601  GTTGATGATCTCAGTAAGAGGCTTCAAGGAATCAAGAAATGGCTCGAATAATTGAC 660
Db      3382  GTTGATGATCTCAGTAAGAGGCTTCAAGGAATCAAGAAATGGCTCGAATAATTGAC 3441
Qy      661  AATCTCAATCAAGCGGTATCAAGAG 685
Db      3442  AATCTCAATCAAGCGGTATCAAGAG 3466

RESULT 5
AX030324      5925 bp      DNA      linear      PAT 16-SEP-2000
LOCUS      Sequence 3 from Patent EP0967279.
ACCESSION      AX030324
VERSION      AX030324.1 GI:10190490
KEYWORDS
SOURCE      Helicobacter pylori
ORGANISM      Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
              Helicobacteriaceae; Helicobacter.
REFERENCE      1 Covacci, A., Rappuoli, R., Bugnoli, M., Telford, J. and Macchia, G.
AUTHORS      Helicobacter pylori cytotoxin useful for vaccines and diagnostics
TITLE      Patent: EP 0967279-A 3 29-DEC-1999;
JOURNAL      CHIRON SPA (IT)

FEATURES
source      1..5925
              /organism="Helicobacter pylori"
              /mol_type="unassigned DNA"
              /db_xref="taxon:210"

ORIGIN
Query Match      100.0%; Score 685; DB 6; Length 5925;
Best Local Similarity 100.0%; Pred. No. 5,2e-118;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  AAAAAATGCAAAAATTAAGATTTTCAGCAAGGTACGCAAGCAAAAAGGACCTTGAAT 60
Db      2782  AAAAAATGCAAAAATTAAGATTTTCAGCAAGGTACGCAAGCAAAAAGGACCTTGAAT 2841
Qy      61  TCCGTTAAAGATGTGATCATCAATCAAAAGGTAAACGATTAAGTTGATTCATCA 120
Db      2842  TCCGTTAAAGATGTGATCATCAATCAAAAGGTAAACGATTAAGTTGATTCATCA 2901
Qy      121  GCGGTATCAAGTGGTAAAGCAACGGGTATTCAGTAGGGTAAAGCAAGGTTAGCCAT 180
Db      2902  GCGGTATCAAGTGGTAAAGCAACGGGTATTCAGTAGGGTAAAGCAAGGTTAGCCAT 2961
Qy      181  CTCAAAAATTTCTCAAAAGAGCAATTTGGCCCAACAGCTCAAAAAATTAAGTCTCAAT 240
Db      2962  CTCAAAAATTTCTCAAAAGAGCAATTTGGCCCAACAGCTCAAAAAATTAAGTCTCAAT 3021
Qy      241  GCTAGAAAAAAATCTGAATATATATATATATATATATATATATATATATATATATAT 300
Db      3022  GCTAGAAAAAAATCTGAATATATATATATATATATATATATATATATATATATATAT 3081
Qy      301  GGTATGGGTATCTCAAGCAAGAGCAACCTTTCTTAAAACTTTTGGACATCAAG 360
Db      3082  GGTATGGGTATCTCAAGCAAGAGCAACCTTTCTTAAAACTTTTGGACATCAAG 3141
Qy      361  AAAGAGTTGAATGCAAACTTGAATTTTCAATTAACAATTAACAATTAATGACTCAAAAC 420
Db      3142  AAAGAGTTGAATGCAAACTTGAATTTTCAATTAACAATTAACAATTAATGACTCAAAAC 3201
Qy      421  GAACCAATTTATGCTAAAGTTAAATTAAGAAAGAGGCAAGGCAAGCACTAGCTGAAGAA 480
Db      3202  GAACCAATTTATGCTAAAGTTAAATTAAGAAAGAGGCAAGGCAAGCACTAGCTGAAGAA 3261

```

```

Qy      481  CCCATTTAGGCTCAAGTTGCTAAAAAGTTAAATGCAAAAATTGACCACTCAATCAATA 540
Db      3262  CCATTTAGGCTCAAGTTGCTAAAAAGTTAAATGCAAAAATTGACCACTCAATCAATA 3321
Qy      541  GCAAGTGGTTGGTGTGTGTAAGGGCAAGCAGCGGGCTTCCCTTGAAGAAAGCATGATAA 600
Db      3322  GCAAGTGGTTGGTGTGTGTAAGGGCAAGCAGCGGGCTTCCCTTGAAGAAAGCATGATAA 3381
Qy      601  GTTGATGATCTCAGTAAGAGGCTTCAAGGAATCAAGAAATGGCTCGAATAATTGAC 660
Db      3382  GTTGATGATCTCAGTAAGAGGCTTCAAGGAATCAAGAAATGGCTCGAATAATTGAC 3441
Qy      661  AATCTCAATCAAGCGGTATCAAGAG 685
Db      3442  AATCTCAATCAAGCGGTATCAAGAG 3466

RESULT 6
BD000835      5925 bp      DNA      linear      PAT 31-JAN-2002
LOCUS      Helicobacter pylori proteins useful for vaccines and diagnostics.
ACCESSION      BD000835
VERSION      BD000835.1 GI:18625394
KEYWORDS      JP 2000350591-A/3.
SOURCE      unidentified
              unidentified
              unidentified
              unclassified.
              1 (bases 1 to 5925)
REFERENCE      1 Covacci, A., Bugnoli, M., Telford, J., Macchia, G. and Rappuoli, R.
AUTHORS      Helicobacter pylori proteins useful for vaccines and diagnostics
TITLE      Patent: JP 2000350591-A 3 19-DEC-2000;
JOURNAL      CHIRON SPA

COMMENT
OS      Unidentified
PN      JP 2000350591-A/3
PD      19-DEC-2000
PF      26-APR-2000 JP 2000126695
PR      02-MAR-1992 IT FI92A000052, 25-JAN-1993 GB PCTEP9300158 PI
AN       ANTONELLO COVACCI, MASSIMO BUGNOLI, JOHN TELFORD, PI GIOVANNI
MA       MACCHIA,
PI       RINO RAPPUOLI
PC      C12N15/09, A61K39/106, A61K39/39, A61K48/00, A61P1/04,
PC      A61P31/04,
PC      A61P35/00, C07K14/195, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC
PC      C12P21/02,
PC      C12Q1/68, G01N33/569//C12N15/09, C12R1.01, (C12N1/21, C12R1.19),
PC      C12R1.01, C12R1.19), C12N15/00, A61K37/02, C12N5/00, (C12N15/00,
CC      Strandedness: Single;
CC      Topology: Linear;
PH      Key
FT      source      1..5925
              location/Qualifiers
              1..5925
              /organism="unidentified"
              /mol_type="genomic DNA"
              /db_xref="taxon:3264"

ORIGIN
Query Match      100.0%; Score 685; DB 6; Length 5925;
Best Local Similarity 100.0%; Pred. No. 5,2e-118;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  AAAAAATGCAAAAATTAAGATTTTCAGCAAGGTACGCAAGCAAAAAGGACCTTGAAT 60
Db      2782  AAAAAATGCAAAAATTAAGATTTTCAGCAAGGTACGCAAGCAAAAAGGACCTTGAAT 2841
Qy      61  TCCGTTAAAGATGTGATCATCAATCAAAAGGTAAACGATTAAGTTGATTCATCA 120
Db      2842  TCCGTTAAAGATGTGATCATCAATCAAAAGGTAAACGATTAAGTTGATTCATCA 2901
Qy      121  GCGGTATCAAGTGGTAAAGCAACGGGTATTCAGTAGGGTAAAGCAAGGTTAGCCAT 180

```

```

Db      2902 GCGGATCAGTGGCTAAAGCAACGGGTGATTCAGTAGGGTAGAGCAAGCGTTAGCGAT 2961
Qy      181 CTCAAAATTTCTCAAGGAGCAATGGCCCAACAAGCTCAAAAAATGAAATCTCAAT 240
Db      2962 CTCAAAATTTCTCAAGGAGCAATGGCCCAACAAGCTCAAAAAATGAAATCTCAAT 3021
Qy      241 GCTAGAAAAAATCTGAATATATATCATCCGTTAGAGTGTGTGAATGGAACCTAGTC 300
Db      3022 GCTAGAAAAAATCTGAATATATATCATCCGTTAGAGTGTGTGAATGGAACCTAGTC 3081
Qy      301 GGTAAATGGTTATCTCAAGCAAGACCAACTCTTTCTAAAAAATTTTGGACATCAAG 360
Db      3082 GGTAAATGGTTATCTCAAGCAAGACCAACTCTTTCTAAAAAATTTTGGACATCAAG 3141
Qy      361 AAGAGGTGAATGCAAAATCTTGAAATTTCAATACATACATTAATGATCTCAAAAC 420
Db      3142 AAGAGGTGAATGCAAAATCTTGAAATTTCAATACATACATTAATGATCTCAAAAC 3201
Qy      421 GAACCAATTTATCTAAAGTTAATTAATAAGAAAGAGGAGAGAGCTAGCTTGAAGAA 480
Db      3202 GAACCAATTTATCTAAAGTTAATTAATAAGAAAGAGGAGAGAGCTAGCTTGAAGAA 3261
Qy      481 CCCATTACGCTCAAGTTCCTAAAGAGTAATGCAAAATTTGACGACTCAATCAATA 540
Db      3262 CCCATTACGCTCAAGTTCCTAAAGAGTAATGCAAAATTTGACGACTCAATCAATA 3321
Qy      541 GCAGAGTGTGGGTGTGTGAGGCAAGCAGCGGCTTCCCTTTGAAAAGGCAATGATAA 600
Db      3322 GCAGAGTGTGGGTGTGTGAGGCAAGCAGCGGCTTCCCTTTGAAAAGGCAATGATAA 3381
Qy      601 GTTATGATCTCAGTAGAGGAGGCTTCAAGGATCAAGAAATTTGCTCAGAAATTTGAC 660
Db      3382 GTTATGATCTCAGTAGAGGAGGCTTCAAGGATCAAGAAATTTGCTCAGAAATTTGAC 3441
Qy      661 AATTCATCAAGCGGTATCAGAG 685
Db      3442 AATTCATCAAGCGGTATCAGAG 3466

RESULT 7
BD000839 5925 bp DNA linear PAT 31-JAN-2002
LOCUS Helicobacter pylori proteins useful for vaccines and diagnostics.
DEFINITION BD000839
ACCESSION BD000839
VERSION BD000839.1 GI:18625398
KEYWORDS JP 2000333686-A/3.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 5925)
AUTHORS Covacci,A., Bugnoli,M., Telford,J., Macchia,G. and Rappuoli,R.
TITLE Helicobacter pylori proteins useful for vaccines and diagnostics
JOURNAL Patent: JP 2000333686-A 3 05-DEC-2000;
CHIRON SPA
COMMENT OS Unidentified
PN JP 2000333686-A/3
PD 05-DEC-2000
PP 26-APR-2000 JP 2000126696
PR 02-MAR-1992 IT FI92A000052,25-JAN-1993 GB PCTEP9300158 PI
ANTONELLO COVACCI,MASSIMO BUGNOLI,JOHN TELFORD, PI GIOVANNI
MACCHIA.
PI RINO RAPPUOLI
PC C12N15/09,A61K38/00,A61K39/106,A61P1/04,C07K14/195, PC
C12N9/80//C12P21/02,
PC (C12N15/09,C12R1:01),C12N15/00,A61K37/02,(C12N15/00,C12R1:01)
CC Strandedness: Single;
CC Topology: Linear;
CC Key Location/Qualifiers
FH key 1..5925
FT source
FEATURES
source Location/Qualifiers
1..5925

```

```

ORIGIN
Query Match 100.0%; Score 685; DB 6; Length 5925;
Best Local Similarity 100.0%; Pred. No. 5,2e-118;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

Qy      1 AAAAAATGCAAAATTAAGGATTTCAAGCAAGTAAAGCAAGCAAAAGCGACCTTGAAAT 60
Db      2782 AAAAAATGCAAAATTAAGGATTTCAAGCAAGTAAAGCAAGCAAAAGCGACCTTGAAAT 2841
Qy      61 TCCGTTAAAGATGTATCATCATCAAAAGGTAAGGATTAAGTTGATATCTCAATCA 120
Db      2842 TCCGTTAAAGATGTATCATCATCAAAAGGTAAGGATTAAGTTGATATCTCAATCA 2901
Qy      121 GCGGTATCAGTGGCTAAAGCAAGGGTATTTCAATAGGGGTAGCAAGCGTTAGCCGAT 180
Db      2902 GCGGTATCAGTGGCTAAAGCAAGGGTATTTCAATAGGGGTAGCAAGCGTTAGCCGAT 2961
Qy      181 CTCAAAATTTCTCAAGGAGCAATGGCCCAACAAGCTCAAAAAATGAAATGCTCAAT 240
Db      2962 CTCAAAATTTCTCAAGGAGCAATGGCCCAACAAGCTCAAAAAATGAAATGCTCAAT 3021
Qy      241 GCTAGAAAAAATCTGAATATATATCATCCGTTAAGAAATGTGTGAATGGAACCTAGTC 300
Db      3022 GCTAGAAAAAATCTGAATATATATCATCCGTTAAGAAATGTGTGAATGGAACCTAGTC 3081
Qy      301 GGTAAATGGTTATCTCAAGCAAGCAACTCTTTCTAAAAAATTTTGGACATCAAG 360
Db      3082 GGTAAATGGTTATCTCAAGCAAGCAACTCTTTCTAAAAAATTTTGGACATCAAG 3141
Qy      361 AAGAGGTGAATGCAAAATCTTGAAATTTCAATACATTAATGATCTCAAAAC 420
Db      3142 AAGAGGTGAATGCAAAATCTTGAAATTTCAATACATTAATGATCTCAAAAC 3201
Qy      421 GAACCAATTTATCTAAAGTTAATTAATAAGAAAGAGGAGAGAGCTAGCTTGAAGAA 480
Db      3202 GAACCAATTTATCTAAAGTTAATTAATAAGAAAGAGGAGAGAGCTAGCTTGAAGAA 3261
Qy      481 CCCATTACGCTCAAGTTCCTAAAGAGTAATGCAAAATTTGACGACTCAATCAATA 540
Db      3262 CCCATTACGCTCAAGTTCCTAAAGAGTAATGCAAAATTTGACGACTCAATCAATA 3321
Qy      541 GCAGAGTGTGGGTGTGTGAGGCAAGCAGCGGCTTCCCTTTGAAAAGGCAATGATAA 600
Db      3322 GCAGAGTGTGGGTGTGTGAGGCAAGCAGCGGCTTCCCTTTGAAAAGGCAATGATAA 3381
Qy      601 GTTATGATCTCAGTAGAGGAGGCTTCAAGGATCAAGAAATTTGCTCAGAAATTTGAC 660
Db      3382 GTTATGATCTCAGTAGAGGAGGCTTCAAGGATCAAGAAATTTGCTCAGAAATTTGAC 3441
Qy      661 AATTCATCAAGCGGTATCAGAG 685
Db      3442 AATTCATCAAGCGGTATCAGAG 3466

RESULT 8
AP282853 20229 bp DNA linear BCT 16-AUG-2000
LOCUS Helicobacter pylori cagI pathogenicity island, complete sequence.
DEFINITION AP282853 U60176
ACCESSION AP282853
VERSION AP282853.1 GI:8843984
KEYWORDS
SOURCE
ORGANISM Helicobacter pylori
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteraceae; Helicobacter.
REFERENCE 1 (bases 1 to 20229)
AUTHORS Censini,S., Lange,C., Xiang,Z., Crabtree,J.E., Ghiara,P.,
Borodovsky,M., Rappuoli,R. and Covacci,A.
TITLE cagI, a pathogenicity island of Helicobacter pylori, encodes type

```

```
I-specific and disease-associated virulence factors
Proc. Natl. Acad. Sci. U.S.A. 93 (25), 14646-14653 (1996)
REFERENCE
PUBMED 8962108
AUTHORS Covacci, A., and Rappuoli, R.
TITLE Tyrosine-phosphorylated bacterial proteins: Trojan horses for the
host cell
JOURNAL J. Exp. Med. 191 (4), 587-592 (2000)
PUBMED 10684850
REFERENCE 3 (bases 1 to 20229)
AUTHORS Censini, S., Rappuoli, R., Lange, C. and Covacci, A.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-1996) Molecular Biology, Chiron-Biocrine, Via
4 (bases 1 to 20229)
REFERENCE 4 (bases 1 to 20229)
AUTHORS Censini, S., Rappuoli, R. and Covacci, A.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-2000) Molecular Biology, Chiron-Biocrine, Via
Florentina 1, Siena, SI 53100, Italy
REMARK Sequence update by submitter
COMMENT On Jun 30, 2000 this sequence version replaced gi:1752680.
FEATURES
Source
1. .20229
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="CCUG 17874; NCTC 11638"
/db_xref="taxon:210"
/note="Type I-specific and disease-associated virulence
factor genes; partial sequence of cag pathogenicity island
presented in AF282852"
1. .1880
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="CCUG 17874; NCTC 11638"
/db_xref="taxon:210"
/insertion_seq="IS605"
/note="Genbank Accession Number U60177"
18351..18610
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="CCUG 17874; NCTC 11638"
/db_xref="taxon:210"
/insertion_seq="IS605"
/note="Relic of IS605"
complement(83..522)
/gene="tnpA"
complement(83..511)
/gene="tnpB"
/note="similar to IS200 transposase encoded by GenBank
Accession Number U60177"
codon_start=1
/transl_table=11
/product="tnpA"
/protein_id="AAC44689.1"
/db_xref="GI:1752683"
/translation="MRKNSPLRGYSITNRSNKKAKHLIVCKYRKLLQGLDNFI
KSVIDEIATQSHPTITNHSIDHILAMQVYIPKSISSITIKIKQITTYRWMDKRF
IPLQKPMKPKPTDPSYCSIGQANETTKAYTENQ"
complement(519..522)
/gene="tnpA"
/note="putative"
570..1864
/gene="tnpB"
570..574
/gene="tnpB"
/note="putative"
581..1864
/gene="tnpB"
/note="similar to IS1341 transposase encoded by GenBank
Accession Number U60177"
codon_start=1
/transl_table=11
/product="tnpB"
/protein_id="AAC44692.1"
/db_xref="GI:1752686"
/translation="MPLTNRIRDPNKKOLNPKIKGLIMGKILSLGGQTLPTGL
SDLPFMINLPLFLMIMGLNEALGKKFNLPMDNINKMFAVLKNGPDSIKMGSAL
VGNQFSSKSDKTNKMSVOYAL"
2143..2455
/gene="cagA"
2143..2146
/gene="cagB"
/note="putative"
2153..2455
/gene="cagB"
/codon_start=1
/transl_table=11
/product="CagB"
/protein_id="AAC44692.1"
/db_xref="GI:1752686"
/translation="MPLISNPFSTSAIKRLLISGLKLNFPNASSLPTINRNKK
FNLKSLDKPVRKFPVPPNKKAKIFPMISPIFCVSSCLFGLMRVVGSIQ"
complement(2459..2462)
/gene="cagC"
complement(2878..3311)
/gene="cagP"
complement(2878..3210)
/gene="cagP"
/note="HP536; cagI, JHP484"
/codon_start=1
/transl_table=11
/product="CagP"
/protein_id="AAP80206.1"
/db_xref="GI:8843986"
/translation="MSKLNKPNLQPKYSFNKKLDKYSLLYRLFNISIVIGLGLFS
VAGAVITLILFLLPFLALIKPSFFYYTTLVLVLSIISKYLLSHAKFTKLLIIM
TOMQWML"
complement(3219..3221)
/gene="cagP"
/note="putative"
complement(3306..3536)
/gene="cagO"
complement(3306..3521)
/gene="cagO"
/codon_start=1
/transl_table=11
/product="CagO"
/protein_id="AAC44694.1"
/db_xref="GI:1752688"
/translation="MPLVVGGLVFFKFLRLPLVYRNVFFRNKTPLEFNRCPLFV
WHKQTNMFWLHYMLVSSAGVPEIWS"
complement(3533..3536)
/gene="cagO"
/note="putative"
3628..4777
/protein_id="AAP80205.1"
/db_xref="GI:8843985"
/translation="MLNLIKRTYIPNNAQOKELISNHCGRVYVYVYELDYRQKQYANG
IKETVPMQKVLTOIKHOEKYHVNLCNSQSIOMALQOALVNDVFSKRAVFKFS
KNNAKOSFAPLPIONIEKTEGTIALPKPREGIKATRELPDNDVFSKRAVFKFS
FCSISYETKEPIPKPTIIKAVGLDGLGTLTLYSDKTPYVPIRYPQVQNFISCIADY
RLSKKVGSNRRKQAKVAKRLHACSNTRDYLHKITSNITNODYQVQNFISCIADY
MRVYKSLANASMGKPLTMKTKAQKAKTLLGIDPPSPSQCSYCPYVGHKSL
ITKFTCHCNITHRHDVYASVIRNYALGMLDDBHKIKIDKSRVGIKTDVATHTDR
IKACGASNSGVISKGNITLDASVGMKQKAGSL"
1881..19352
/note="cagI pathogenicity island"
complement(2065..2462)
/gene="cagO"
complement(2065..2445)
/gene="cagO"
/note="HP535; cagI4; JHP483"
complement(2065..2445)
/gene="cagO"
/note="HP535; cagI4; JHP483"
/codon_start=1
/transl_table=11
/product="CagO"
/protein_id="AAC44691.1"
/db_xref="GI:1752685"
/translation="MPLTNRIRDPNKKOLNPKIKGLIMGKILSLGGQTLPTGL
SDLPFMINLPLFLMIMGLNEALGKKFNLPMDNINKMFAVLKNGPDSIKMGSAL
VGNQFSSKSDKTNKMSVOYAL"
2143..2455
/gene="cagA"
2143..2146
/gene="cagB"
/note="putative"
2153..2455
/gene="cagB"
/codon_start=1
/transl_table=11
/product="CagB"
/protein_id="AAC44692.1"
/db_xref="GI:1752686"
/translation="MPLISNPFSTSAIKRLLISGLKLNFPNASSLPTINRNKK
FNLKSLDKPVRKFPVPPNKKAKIFPMISPIFCVSSCLFGLMRVVGSIQ"
complement(2459..2462)
/gene="cagC"
complement(2878..3311)
/gene="cagP"
complement(2878..3210)
/gene="cagP"
/note="HP536; cagI, JHP484"
/codon_start=1
/transl_table=11
/product="CagP"
/protein_id="AAP80206.1"
/db_xref="GI:8843986"
/translation="MSKLNKPNLQPKYSFNKKLDKYSLLYRLFNISIVIGLGLFS
VAGAVITLILFLLPFLALIKPSFFYYTTLVLVLSIISKYLLSHAKFTKLLIIM
TOMQWML"
complement(3219..3221)
/gene="cagP"
/note="putative"
complement(3306..3536)
/gene="cagO"
complement(3306..3521)
/gene="cagO"
/codon_start=1
/transl_table=11
/product="CagO"
/protein_id="AAC44694.1"
/db_xref="GI:1752688"
/translation="MPLVVGGLVFFKFLRLPLVYRNVFFRNKTPLEFNRCPLFV
WHKQTNMFWLHYMLVSSAGVPEIWS"
complement(3533..3536)
/gene="cagO"
/note="putative"
3628..4777
```



```

RBS      /gene="cagm"
          3628..3631
          /gene="cagm"
          /note="putative"
          3647..4777
CDS
          /gene="cagm"
          /note="HP537; cag16; JHP485"
          /codon_start=1
          /transl_table=11
          /product="CagM"
          /protein_id="AAC44695.1"
          /db_xref="GI:1752689"
          /translation="MLAKIVSSILVAFGLSANVQPSFPEIKKEQEVAKEDAL
          KARKKILNTHDFLEDLIPRKOKIKELMDHRAVYLSDLKMKKKKBALEKTRGTL
          TAKSKAYGDLKONPLKYLKLPNPYAVVLAQVPTTKEDKRLSYTPQVTSIF
          KRTATATTKDRAQLLQMGVSLDBEQKASRLALSTKQIBESYNNISMLSRKELD
          NIDYVLQLENNKFDKAKDIAQKATWLLPNSERLAFSMAIDIKENKYLKGYAFNL
          LKRVKQDLENTLTKNFTNOKLSFAOKOKLCLVDSFNDTQSKKSLIKKTNYNIF
          VSDPMMSDKTVMQKHYKLPFFKTVSAVRNNVAKNPF"
          4781..5712
          /gene="cagm"
          4781..4785
          /gene="cagm"
          /note="putative"
          4792..5712
          /gene="cagm"
          /note="HP538; cag17; JHP486"
          /codon_start=1
          /transl_table=11
          /product="CagN"
          /protein_id="AAF80207.2"
          /db_xref="GI:9832971"
          /translation="MKSIRPKKLSVALYSLVYGGIANTALLPSEYKELVALGPKK
          IKTLQYRHDKETITKSEKFAINALREKLNDRARLEQIKNTAEKKNSSVQKA
          AKRKLQELNEIANPLNDPNNGSSSTETKSNDRDDEIMINKVNSFYKPAFLVDP
          EWRTPEIBIINECIIISNDYGLRKLIKIDIKQKILAPLEKIOEITBNNKFSRO
          HLSGLKLTNNNNRTFLASDCAICERKKMEQENNYQDTTASRFGTTPTKEMNAK
          DTAFNNRSKSELPSNVINOISIAHGK"
          complement(5794..6507)
          /gene="cagL"
          complement(5794..6507)
gene
CDS
Query Match      100.0%; Score 685; DB 1; Length 20229;
Best Local Similarity 100.0%; Pred. No.3.3e-118;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 421 GAACCCATTATCTAAAGTTAATAAAGAAAGACAGGCGACAGCTAGCTTGAAGA 480
Db 17505 GAACCCATTATCTAAAGTTAATAAAGAAAGACAGGCGACAGCTAGCTTGAAGA 17564
Qy 481 CCATTTAGCGCTCAAGTTGCTAAAGTAATGCAAAAATTGACCGACTCAATCAATA 540
Db 17565 CCATTTAGCGCTCAAGTTGCTAAAGTAATGCAAAAATTGACCGACTCAATCAATA 17624
Qy 541 GCAAGTGGTTGGGTGTGTAAGGCAAGACGCGGCTTCCCTTTGAAAAGCATGATAA 600
Db 17625 GCAAGTGGTTGGGTGTGTAAGGCAAGACGCGGCTTCCCTTTGAAAAGCATGATAA 17684
Qy 601 GTTAGATCTCAGTAGGAGGCTTTCAAGATCAAGAAATGGCTCAGAAATGAC 660
Db 17685 GTTAGATCTCAGTAGGAGGCTTTCAAGATCAAGAAATGGCTCAGAAATGAC 17744
Qy 661 AATCTCAATCAAGCGGTATCAGAG 685
Db 17745 AATCTCAATCAAGCGGTATCAGAG 17769

RESULT 9
LOCUS AB015416 3444 bp DNA linear BCT 13-JUN-2001
DEFINITION Helicobacter pylori caga gene, complete cds, strain NCTC11637.
ACCESSION AB015416
VERSION AB015416.1 GI:12225025
KEYWORDS caga.
SOURCE Helicobacter pylori
ORGANISM Helicobacter pylori
REFERENCE 1 (sites)
AUTHORS Hoshino,F.B., Katayama,K., Watanabe,K., Takahashi,S., Uchimura,H.
and Ando,T.
TITLE Heterogeneity found in the caga gene of Helicobacter pylori from
JOURNAL Japanese and non-Japanese isolates
PUBMED 11573724
REFERENCE 2 (bases 1 to 3444)
AUTHORS Hoshino,F.B.
TITLE Direct Submision
JOURNAL Submitted (09-JUN-1998) Fumimori B Hoshino, Biomedical
Laboratories, Inc, Basic Research Division, 1361-1 Maroba, Kawagoe,
Saitama 350-1101, Japan (E-mail:boone3@alk.co.jp,
Tel:81-492-32-0440, Fax:81-492-32-5480)
FEATURES
source
1..3444
location/Qualifiers
1..3444
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="NCTC 11637"
/db_xref="taxon:210"
1..3444
/gene="caga"
1..3444
/gene="caga"
1..3444
/codon_start=1
/transl_table=11
/product="Caga"
/protein_id="BAB20926.1"
/db_xref="GI:12225026"
/translation="MTNETIDQOPTEAFNPQOFINNLOVAFLKVDNAVASVDPDQK
SIVDNDRNOAFEGISQLEBEYSNKAIKPTKKQVSEFINKSNLIMKNDLIDV
ESSTKSPQKPDQVRIETSVSHQNDPSKNTNISRPMENITOPPLIDDKKAEPL
KSAKSPKGIITIGNOIRTDQKFMGVFDSLRQBAEKNGSPGTGWDLIDFLSTFDK
KQSDVKAINQEPVPHQPDIAITTTDIDQLPEPADLDERGAFKFTLGDMEMLD
VEGYADIDPNYKFNQLIHNNALSVIAGSHNGIEPERVSLIYGGNGGPGARHDMNAT
VGYXDOQGNANVATIIIVHMKNGSGLVLAGSKGJNBPSEFYLYKEDOLGSRALSOE
IONKIDFMEFLAONNAKLDNLSEKKEKFTREIDPOFSKAYVLDALGNDLITPVSK
DTKSALETTERGNDISYTLKDYGRKADKLDREKNTLQSLGHDGVMFVDSNFKY
TNASINPKKGYGTVNGVSHLEVGKRAIATFPLDNLNATISFVRNLEDDLTGSL
FOEANKLIDFLSSNKLVLGKTLNPNNAVADAKNTGVDEVKQAKOLEKSLRRREHL

```

EEVEKTLSEKSGKRRKREKRAQANSQDEIPLINKEANDRAIAVAYONLKGIKRE
LSDLENNVNLKDFPKSDEKNGKPKFSVAEETTLKALKGSKVLGINPEMISKE
NMINAIEFKNGKNDKFSKTOAKSDIENSVDVINOQTYDVKDMLNOAVSAKAG
DESEVEONADLNRKSKQLAQOAKRRESINARKESEIYOSVNGVNGTLVNGLSOA
EATTLSEKESDIKEELANKLGNFNNNNNNNNKNEEIIYAKNKKKGAQAALEETIYO
VAKVNAKIDRLNOLASGLSVGQAAGFELKRDVDLSEKGLSRNOELAKIDININ
QAVSEKAKGFTNLSQTLIDKLKOSTKHNEMNIMWESAKKVPASLSKIDNATNSHR
INSYKNGALNEKATGMLTOKNPEMLKLVNDKIVAHVNGSVPLSEYDKIGNOKMKD
YSDSKFSTLNNAVKDINSGLTQPLTNASTASYCLARENAHEGIIQNVTKGGRQK
S"

ORIGIN

Query Match Best Local Similarity 99.9%; Score 683.4; DB 1; Length 3446;
Matches 684; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAATGCAAAATAAGATTTCAGCAAGGTAAAGCAAGCAAAAGCCAGCTTGAATAAT 60
Db 2248 AAAAAATGCAAAATAAGATTTCAGCAAGGTAAAGCAAGCAAAAGCCAGCTTGAATAAT 2307
Qy 61 TCCGTTAAAGATGTATCATCATCAAAAGGTAAAGCAAGGTAAAGTATATCTCAATCA 120
Db 2308 TCCGTTAAAGATGTATCATCATCAAAAGGTAAAGCAAGGTAAAGTATATCTCAATCA 120
Qy 121 GCGGTATCAGTGTAAAGCAAGCGGTGATTCAGTACGGTAAAGCAAGCGGTGATTCAGT 180
Db 2368 GCGGTATCAGTGTAAAGCAAGCGGTGATTCAGTACGGTAAAGCAAGCGGTGATTCAGT 180
Qy 181 CTCAAAAATTTCTCAAGAGCAATTCAGCAAGGTGATTCAGTACGGTAAAGCAAGCGGTGATTCAGT 240
Db 2428 CTCAAAAATTTCTCAAGAGCAATTCAGCAAGGTGATTCAGTACGGTAAAGCAAGCGGTGATTCAGT 240
Qy 241 GCTAGAAAAAATTCGAAATATATCAATCCGTAAAGAAAGGTGTATGATGAAAGCTTAACT 2487
Db 2488 GCTAGAAAAAATTCGAAATATATCAATCCGTAAAGAAAGGTGTATGATGAAAGCTTAACT 2487
Qy 301 GGTATAGGTATATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 360
Db 2548 GGTATAGGTATATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 360
Qy 361 AAGAGTGAATGCAAAATTCGAAATATATCAATCCGTAAAGAAAGGTGTATGATGAAAGCTTAACT 420
Db 2608 AAGAGTGAATGCAAAATTCGAAATATATCAATCCGTAAAGAAAGGTGTATGATGAAAGCTTAACT 420
Qy 421 GAACCAATTTATGCTAAAGTAAATTAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2667
Db 2668 GAACCAATTTATGCTAAAGTAAATTAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2667
Qy 481 CCCATTAGAGCTCAAGTGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 540
Db 2728 CCCATTAGAGCTCAAGTGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 540
Qy 541 GCAAGTGTGTTGGTGTGTTGTTAGGGCAAGCAAGCGGTGTTCCCTTTGAAAGGAGTATATAA 600
Db 2788 GCAAGTGTGTTGGTGTGTTGTTAGGGCAAGCAAGCGGTGTTCCCTTTGAAAGGAGTATATAA 600
Qy 601 GTTGATGATCTCAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2847
Db 2848 GTTGATGATCTCAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2847
Qy 661 AATTCATCATCAAGCGGTATCAAGAG 685
Db 2908 AATTCATCATCAAGCGGTATCAAGAG 2932

RESULT 10
DO011619 3426 bp DNA linear BCT 01-JUL-2005
LOCUS Helicobacter pylori strain 2808 Caga gene, partial cds.
DEFINITION DO011619
ACCESSION DO011619
VERSION DO011619.1 GI:62871422
KEYWORDS Helicobacter pylori
SOURCE

ORGANISM

Helicobacter pylori

Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteraceae; Helicobacter.

REFERENCE

1 (bases 1 to 3426)

AUTHORS

Brandt, S., Kwok, T., Hartig, R., Konig, W., and Backert, S.

TITLE

NE-(kappa)B activation and potentiation of proinflammatory

JOURNAL

responses by the Helicobacter pylori Caga protein

PUBMED

15972330

REFERENCE

2 (bases 1 to 3426)

AUTHORS

Backert, S.

TITLE

Direct Submersion

JOURNAL

Submitted (18-APR-2005) Medical Microbiology, Otto von Guericke

FEATURES

Location/Qualifiers

source

1. 3426

CDS

/organism="Helicobacter pylori"

FEATURES

/mol_type="genomic DNA"

FEATURES

/strain="2808"

FEATURES

/db_xref="taxon:210"

FEATURES

1. 3426

FEATURES

/codon_start=1

FEATURES

/product="Caga"

FEATURES

/protein_id="AAV18597.1"

FEATURES

/db_xref="gi:62871423.1"

FEATURES

/translation="MTNETTDDQPTAEAFNPQGFNNLQVAFKLVNVAASVDPDQK

FEATURES

PIYDKNDNDNQAPEKISQIAEERFANNAIKNPKNQVYFSDIFDKSNDLINDKNDIDV

FEATURES

ESSTSPKREDDQRTKFTQWVSHQNDPSKINRPMENIIOPPIDDKERAEFL

FEATURES

KAKQSPKGLIIIGRQRTDQKRWGVDESLKQERKNGKGGTGMGLDIPLSFIFDK

FEATURES

YSCVDLDIDPRTKTNQLIHNNALSSVLMGSHNGIPEFVSLIKVGGGPEADHMDNP

FEATURES

LYITKTRQCGSLINLVHMKNSGLVIGGKSKSPFNEIEDFQDSFAYLDLLENDHIAFVSKDKK

FEATURES

HAALVAFEGEILSYTLQDKRKQDKALBREKRTTLQGNLKHGWFVNVSNPKYTA

FEATURES

SQSPKVGATNGVSHLEBANISKVAVNLPDJNLALITSYLRDLDDKLRAGSLSPQK

FEATURES

YMKLLQDLPLNSTKEIYERKALNFNTYVAENKNTGANDVEYKAQDLDEKSLRGRHLEK

FEATURES

YKLENINRNSKDPKSFDEFKNGKMDFAFRTTLKPLRFRFPYQSRMDPFVENVLNTA

FEATURES

LEERFNGKNDPSKTYQAKSDLENSIKDYIINQKITDVRNMLQAVSLAKGIDRSPV

FEATURES

EGALNDLNRKSKQLAQOAKRRESINARKESEIYAKVNNKKTGOVSLSEPIYTOYAKV

FEATURES

TRNFSDIKEELNGLKUNNNNNNNNNKNEEIIYAKNKKKGAQAALEETIYOYAKV

FEATURES

NAKIDRLNOLASGLNVGQAVGPELRKHDKVDLSKGLSRNOELAKIDININQAVR

FEATURES

AKAGPFGNLEQRTIGNLKDSAHNPMNLWESAKKVPISLSKIDNATNSHRINSY

FEATURES

KNGALNEKATGMLTOKNPEMLKLVNDKIVAHVNGSVPLSEYDKIGNOKMKD

FEATURES

YSDSKFSTLNNAVKDINSGLTQPLTNASTASYCLARENAHEGIIQNVTKGGRQK

FEATURES

S"

FEATURES

Qy 87.9%; Score 601.8; DB 1; Length 3426;
Best Local Similarity 92.4%; Pred. No. 2e-102;
Matches 633; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

ORIGIN

Qy 1 AAAAAATGCAAAATAAGATTTCAGCAAGGTAAAGCAAGCAAAAGCCAGCTTGAATAAT 60
Db 2233 AAAAAATGCAAAATAAGATTTCAGCAAGGTAAAGCAAGCAAAAGCCAGCTTGAATAAT 2292
Qy 61 TCCGTTAAAGATGTATCATCATCAAAAGGTAAAGCAAGGTAAAGTATATCTCAATCA 120
Db 2293 TCCGTTAAAGATGTATCATCATCAAAAGGTAAAGCAAGGTAAAGTATATCTCAATCA 120
Qy 121 GCGGTATCAGTGTAAAGCAAGCGGTGATTCAGTACGGTAAAGCAAGCGGTGATTCAGT 180
Db 2353 GCGGTATCAGTGTAAAGCAAGCGGTGATTCAGTACGGTAAAGCAAGCGGTGATTCAGT 180
Qy 181 CTCAAAAATTTCTCAAGAGCAATTCAGCAAGGTGATTCAGTACGGTAAAGCAAGCGGTGATTCAGT 240
Db 2413 CTCAAAAATTTCTCAAGAGCAATTCAGTACCAAGCTCAAAAAAATGAAGATTTCATAT 2472
Qy 241 GCTAGAAAAAATTCGAAATATATCAATCCGTAAAGATGTGTAATGAAAGCCCTAGTAC 300
Db 2473 ACTGAAAAAATTTCTGAACCTATCCGTAAAGATGTGTAATGAAAGCCCTAGTAC 2532
Qy 301 GGTATAGGTATATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 360

```

|||||
Db 2533 GGTAAATGGGTTATCTGGAAATAGAGCCACAGCTCTCAACAAAATTTTTCGATATCAAG 2592
Qy 361 AAAGAGTGAATCCAAAACCTTGCAAAATTTCAATTAACAATAACAATAATGACCTCAAAAC 420
Db 2593 AAAGAAATGATGAGAACTTAAATAATTTCAATTAACAATAACAATAATGAGCTCAAAAC 2652
Qy 421 GAACCAATTTATCTAAAGTTAAATTAATAAAGAAAGAGGAGGAGAGCTAGCTTGAAGAA 480
Db 2653 GAACCAATTTATCTAAAGTTAAATTAATAAAGAAAGAGGAGGAGCTAGCTTGAAGAA 2712
Qy 481 CCCATTTAGCGCTCAAGTTGCTTAAATAAAGTAATGCAAAATTTGACGACTCAATCAATA 540
Db 2713 CCCATTTATCTCAAGTTGCTTAAATAAAGTAATGCAAAATTTGACGACTCAATCAATA 2772
Qy 541 GCAAGTGGTTGGGTGTGTGAGGCAAGCAGCGGCTTCCCTTTGAAAAGGCTGATTA 600
Db 2773 GCAAGTGGTTGGGTGTGTGAGGCAAGCAGCGGCTTCCCTTTGAAAAGGCTGATTA 2832
Qy 601 GTTGATGATCTCAGTAAGGAGGCTTTCAGGAATCAAGATTTGGCTCAGAAAATTTGAC 660
Db 2833 GTTGATGATCTCAGTAAGGAGGCTTTCAGGAATCAAGATTTGGCTCAGAAAATTTGAC 2892
Qy 661 AATCTCAATCAAGCGGTATCAGAG 685
Db 2893 AATCTCAATCAAGCGGTATCAGAG 2917

RESULT 11
ABI10963 978 bp DNA linear BCT 23-SEP-2004
LOCUS Helicobacter pylori caga gene for Caga, partial cds, strain:
DEFINITION China:079a.
ACCESSION ABI10963
VERSION ABI10963.1 GI:31339271
KEYWORDS Helicobacter pylori
SOURCE Helicobacter pylori
ORGANISM Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteriaceae; Helicobacter.
REFERENCE
AUTHORS Zhou, W., Yamazaki, S., Yamakawa, A., Ohtani, M., Ito, Y., Keida, Y.,
Higashi, H., Hatakeyama, M., Si, D., and Azuma, T.
TITLE The diversity of vacA and caga genes of Helicobacter pylori in East
JOURNAL Asia
PUBMED FEMS Immunol. Med. Microbiol. 40 (1), 81-87 (2004)
AUTHORS 14734191
2 (bases 1 to 978)
Zhou, W., Yamazaki, S., Yamakawa, A. and Azuma, T.
Direct Submission
Submitted (29-MAY-2003) Takeshi Azuma, Fukui Medical University,
Second Department of Internal Medicine, Matsuoka-cho, Yoshida-gun,
Fukui 910-1193, Japan (E-mail: azuma@fmsara.fukui-med.ac.jp,
Tel: 81-776-61-8351, Fax: 81-776-61-8110)
FEATURES
source
1. 978
Location/Qualifiers
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="China:079a"
/db_xref="taxon:210"
/country="China"
/note="Isolated from a patient with duodenal ulcer"
1. 978
/gene="caga"
<1. >978
/gene="caga"
/codon_start=1
/transl_table=11
/product="Caga"
/protein_id="BAC71006.1"
/db_xref="GI:31339272"
/translation="LKDPSKSFDEFKNGKNDPSKTEETLKALXDSVKDGLNPWMIS
KVENLTALNEFKNGKNDPSKVTQAKSLDLSINDVIINQKITDKVDNINQAVSIAR

```

ANGPSGVEOVLADLKNFSKGOLAOQAKNEDPNTKNSLQSYKNGVNGTLVNGSL
 SOAEATLTSKNSDIDKEINERKPNRNNRNGLNKEPIYAAVKKRGEVASLEPI
 YTOVAVNAKIDRLNOLASGFGVGAQFPLKRDVGLSKGLSREQLKOKID
 NINQAVSEAKCHFDNLVOMIDELDSTKQNVNIMVESAKKVPASLSAK"

Query Match 87.2%; Score 597; DB 1; Length 978;
 Best Local Similarity 92.0%; Pred. No. 2.6e-101;
 Matches 630; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```

Qy 1 AAAAATGGCAAAATATAGGATTTTCAGCAAGTAAGCAAGCAAAAGCGACCTTGAAAT 60
Db 169 AAAAATGGCAAAATATAGGATTTTCAGCAAGTAAGCAAGCAAAAGCGACCTTGAAAT 228
Qy 61 TCGTTAAAGATGTGATCATCATCAAAAGTAACGATAAAGTTGATATCTCAATCAA 120
Db 229 TCGATTAAGAGTGTATCATCATCAAAAGTAACGATAAAGTTGATATCTCAATCAA 288
Qy 121 GCGGTATCAGTGGCTTAAAGCAACGGTGTATTCATAGGTGAGCAAGCTTACGCCAT 180
Db 289 GCGGTATCAGTGGCTTAAAGCAACGGTGTATTCATAGGTGAGCAAGCTTACGCCAT 348
Qy 181 CTCAAAAATTTCTCAAAGGAGCAATTTGGCCCAAGAGCTCAAAAAATGAAGTCTCAAT 240
Db 349 CTCAGAAATTTCTCAAAGGAGCAATTTGGCTCAACAGCTCAAAAAATGAAGTCTCAAT 408
Qy 241 GCTAGAAAAAAATCTGAATATATATCATCCGTTAAGAAATGTGTGATAGAAACCTAGTC 300
Db 409 ACTGAAAAAAATCTGAATATATATCATCCGTTAAGAAATGTGTGATAGAAACCTAGTC 468
Qy 301 GGTATATGGTTATCTCAGACAGACACCACTCTTTTAAAACTTTTGCGACATCAAG 360
Db 469 GGTATATGGTTATCTCAGACAGACACCACTCTTTTAAAACTTTTGCGATATCAAG 528
Qy 361 AAAGAGTGAATGCAAACTTGAAATTTCAATATCAATTAACAATAAGTAAGTAAGTAAGTAAG 420
Db 529 AAAGATTCATGAAATTTTAAATTTTCAACAAATTAACAATAAGTAAGTAAGTAAGTAAGTAAG 588
Qy 421 GAACCAATTTATCTTAAAGTTATTAATAAAGAAAGAGGAGGAGCAAGCTAGCTTGAGAA 480
Db 589 GAACCAATTTATCTTAAAGTTATTAATAAAGAAAGAGGAGGAGCTAGCTTGAGAA 648
Qy 481 CCCATTTAGCGCTCAGTGGCTTAAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 540
Db 649 CCCATTTATCTCAAGTTGCTTAAATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 708
Qy 541 GCAAGTGGTTGGGTGTGTGAGGCAAGCAGCGGCTTCCCTTTGAAAAGGATGATTA 600
Db 709 GCAAGTGGTTGGGTGTGTGAGGCAAGCAGCGGCTTCCCTTTGAAAAGGATGATTA 768
Qy 601 GTTGATGATCTCAGTAAGGAGGCTTTCAGGAATCAAGATTTGGCTCAGAAAATTTGAC 660
Db 769 GTTGATGATCTCAGTAAGGAGGCTTTCAGGAATCAAGATTTGGCTCAGAAAATTTGAC 828
Qy 661 AATCTCAATCAAGCGGTATCAGAG 685
Db 829 AATCTCAATCAAGCGGTATCAGAG 853

RESULT 12
DQ067454 3693 bp DNA linear BCT 15-JUN-2005
LOCUS Helicobacter pylori strain 7bgs cytotoxin-associated protein A
DEFINITION (caga) gene, complete cds.
ACCESSION DQ067454
VERSION DQ067454.1 GI:67483340
KEYWORDS Helicobacter pylori
SOURCE Helicobacter pylori
ORGANISM Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteriaceae; Helicobacter.
REFERENCE
AUTHORS Kim, S.Y., Blaser, M.J., Lee, Y.C. and Pillinger, M.H.

```


gene
 CDS
 complement (2046. .3614)
 /gene="cag8"
 /locus_tag="HP0528"
 /complement (2046. .3614)
 /gene="cag8"
 /locus_tag="HP0528"
 /note="cag-X; jhp0477; orf15; vlrB9-like protein"
 /codon_start=1
 /transl_table=11
 /product="Cag8"
 /protein_id="AA03891.1"
 /db_xref="GI:37811819"
 /translation="MQAPPKIIVNFCGYFLSTIEAVALDINRGRKVVNK
 KIAYLDEKPTITWLDNVTYQLKEDFTISITTFGNGMSIVNSNHIIFOPKSV
 AKNLMEKAVNFMALTRDYOEFLTKTKLIVDPDPKLEBQKALEKEKAEQAK
 AQDKKEKKEERAKNRANLENTNMSNPONLNKNLSLKKQRENLDMERLE
 DMQEOQAANAKOIEELNKKOABETIKORAKDKISIKTDKSKSPEDNSIELSPD
 WRTNIVRTNKALYOFILRIKADONFASAVLYTKLEYPOHREVSITSEELKREBAK
 RKELIKQENLNTATINRYMASNEQITNKEIRKOKTILDDAKALETOYVHAL
 KRPVPRNTNYTOAPERSGHIMPSSIPODGTTFYGFNRTILOPMIFVOPDGKLSM
 TPAIDPNTNSGLRWYRVEIAEKFKLKDVALVTVINKYGNPLTNNYNIKNYGE
 LERVIKCLPLVRDK"
 complement (3667. .5265)
 /gene="cag9"
 /locus_tag="HP0529"
 /complement (3667. .5265)
 /gene="cag9"
 /locus_tag="HP0529"
 /note="cag-W; jhp0478; orf16; vlrB8-like protein"
 /codon_start=1
 /transl_table=11
 /product="Cag9"
 /protein_id="AA03892.1"
 /db_xref="GI:37811820"
 /translation="MKRTFLITLISFPLIVPMKRIADIPVSNKIYQTVCHLIPS
 VYLTKSPADILITGYSIGASGSGSGGCVIEALSTPLATSLASNLKYNLTGEP
 WGSANVAITAIQGFALTPSSGQNFQWNLINKNIDVMSVLDNLSNKIQFTTGGV
 EBNVKNLILQIIGSTTAQNTITADGLIWLIGKFTANKLONNTTAMLAFALESV
 VNGADAAYIPAYGVNLPDIIIGQGSYLDVFSYLYIVFVIGFVFLISFMKLRDISNGIO
 INIGEYMRVFGTLPRKAMVSFIAYAGGYLYKISYIYFGLAGAFGINOYLPNAD
 IYLVNTVNSILPAVAVPNSNNGNAPSLLQGIQVAGISLPAIFMOTYITMRISTVVK
 PLIAGPSTIVPIAVCLVDPFKDSMNILWFINNLPILVLAIPILFVLAFLA
 FNLTTPVAIONIGGIGIDSTIASLITLFLKGFIEITLIESVNAIVNTIFSSVSM
 DGSRRDERDALMVGVSFMFKG"
 complement (5270. .6028)
 /gene="cag10"
 /locus_tag="HP0530"
 /complement (5270. .6028)
 /gene="cag10"
 /locus_tag="HP0530"
 /note="cag-V; jhp0479; orf17"
 /codon_start=1
 /transl_table=11
 /product="Cag10"
 /protein_id="AA03893.1"
 /db_xref="GI:37811821"
 /translation="MLGKNKEVLIDENLVGVIALDRKLAKNKARTPRKAFYLSMA
 LNAVATYSIVMPLPKETDIFVYGIIDRYGEEKIVKRSAPROIVNSBAVVSATSKPV
 SLFGYSKSLADRKDOIMQCVDSROQAMFENRITROPDYKRAEATISSINQRE
 KYKNSPLRLTFITIKITPDMENTEYITTKQVITYIDFAGNSSQENLITNPGFK
 VDIQITDLOQOYVSEIRKIKEVESKNKALKO"
 6414. .7064
 /locus_tag="HP0531"
 /gene="cag11"

CDS
 6414. .7064
 /gene="cag11"
 /locus_tag="HP0531"
 /note="cag-U; jhp0480; orf18"
 /codon_start=1
 /transl_table=11
 /product="Cag11"
 /protein_id="AA03894.1"
 /db_xref="GI:37811822"
 /translation="MNDTYEHGNSLNAPSPNSQNDLNLNLDSLVPKSGIGQRF
 HEALNQEELKNILIEIEKLPQEKYEILMOGAKORIMEAVASFLGIGLEHLL
 GLCMGIFVLPAIYVFLPRTSKNLEVESLTKKLQYRYVAFGAVLFPLETTIRS
 IYELVIGISGNDVLPFLKNICITIGMGIYLYKIKVGIQNFINGLFTSKQGO"
 7077. .7919
 /gene="cag12"
 /locus_tag="HP0532"
 /7077. .7919
 /gene="cag12"
 /locus_tag="HP0532"
 /note="cagT; jhp0481; orf19; vlrB7-like protein"
 /codon_start=1
 /transl_table=11
 /product="Cag12"
 /protein_id="AA03895.1"
 /db_xref="GI:37811823"
 /translation="MKLASVLIQATITCLILSACSNAKVVQKNHVTYPVNELI
 EKYSSIPLNDKIKDIPFWQVQLPNYKDYLLDNQVYVLPKLVHHSKITTLIGDANKI
 LQRYVFOANGARSDIDFLOPTLNQGVMIASVNDNPNSEKRPQTFEDVQSQPM
 LGANTKNHIGVSGANNKRVNVEVAREKAOLEKINQYKTLIDQKOEYTRKNOR
 ELLETLNRAGVQNMQNVISSEIFPNGNLNNQAKEEVREKLQERENBYLNRSL
 LSGK"
 complement (8108. .8707)
 /gene="cag13"
 /locus_tag="HP0534"
 /complement (8108. .8707)
 /gene="cag13"
 /locus_tag="HP0534"
 /note="cagS; jhp0482; orf20"
 /codon_start=1
 /transl_table=11
 /product="Cag13"
 /protein_id="AA03896.1"
 /db_xref="GI:37811824"
 /translation="MSNNMRKLFSMIAOSKQKEKLIESLOENELNTEDEKKIIDQI
 KIMHPFQOMHTNKALDKVLRNWKDRAVYKSIQVDFKVVYOLLESSEIMELHAI
 AKNPFLSKFSDSLIGFLPFPSFQVMEKKSIREMSQIELYGTKLPLKLFWTDE
 EVNFPANLKTIBOYNDVYRDILMFEDLEKMKKGVQNA"
 complement (9148. .9528)
 /gene="cag14"
 /locus_tag="HP0535"
 /complement (9148. .9528)
 /gene="cag14"
 /locus_tag="HP0535"
 /note="cagQ; jhp0483"
 /codon_start=1
 /transl_table=11
 /product="Cag14"
 /protein_id="AA03897.1"
 /db_xref="GI:37811825"
 /translation="WLPKTRIRDPNQELTOPKIKGLMGKIIASLIGGNTNLPGL
 SSDPSMTLNLFLFLMLMGLENLAGKKFNLPMONIKFMALVKNQFDSIKMGSAL
 VNGFGSSKSDKTANKMSVSGVRL"
 complement (9955. .10143)
 /gene="cag15"
 /locus_tag="HP0536"
 /complement (9955. .10143)
 /gene="cag15"
 /locus_tag="HP0536"
 /note="cagP; jhp0484; truncated due to frameshift"
 /codon_start=1
 /transl_table=11
 /product="truncated Cag15"
 /protein_id="AA03898.1"

gene

/db_xref="GI:37811826"
/translation="MILVYPLFLPLILIKSPFYTYTLLLVLSLISIKRYLLSHA
NFTMKLITMTOMWML"
gene
10724..11854
/locus_tag="HP0537"
10724..11854
/gene="cag16"
/locus_tag="HP0537"

Query Match
Best Local Similarity 94.9%; Score 589.2; DB 1; Length 26663;
Matches 609; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 1 AAAAATGCAAAAATAAGATTTCAGACAGTAAACGCAAGCAAAAGCGACCTTGAATAAT 60
Db 24161 AAAAATGCAAAAATAAGATTTCAGACAGTAAACGCAAGCAAAAGCGACCTTGAATAAT 24220
Qy 61 TCCGTTAAAGATGTATCATCATCAATCAAAAGTAAACGATTAAGTTATCTCAATCA 120
Db 24221 TCCATTAAGATGTATCATCATCAATCAAAAGTAAACGATTAAGTTATCTCAATCT 24280
Qy 121 GCGGTATCATGTGCTTAAGCAACGGGTGATTTCACTAGAGTAAAGCAACGGTACCGCAT 180
Db 24281 GCTGTATCATGTGCTTAAGCAACGGGTGATTTCACTAGAGTAAAGCAACGGTACCGCAT 24340
Qy 181 CTCMAAAATTTCTCAAAAGACCAATTTGGCCCAACAGCTCAAAAATGAAGTCTCAAT 240
Db 24341 CTCMAAAATTTCTCAAAAGACCAATTTGGCTCAAAAGCTCAAAAATGAAGTCTCAAT 24400
Qy 241 GCTAGAAAATAATCTGAATATATCAATCCGTTAAGATGTTGTAATGAAACCTTAAT 300
Db 24401 ACTGAAAAATAATCTGAATATATCAATCCGTTAAGATGTTGTAATGAAACCTTAAT 24460
Qy 301 GGTATAGGTATCTCAAGACGAGCAACCACTCTTTCTTAAAACTTTTGGACATCAAG 360
Db 24461 GGTATAGGTATCTCAAGACGAGCAACCACTCTTTCTTAAAACTTTTGGACATCAAG 24520
Qy 361 AAAGAGTGAATGCAAAATCTGAAATTTTCAATTAACAATTAATGAATGAATCAAAAC 420
Db 24521 AAAGAGTGAATGCAAAATTTGGGAAATTTCAATTAACAATTAATGAATGAATCAAAAC 24580
Qy 421 GAACCAATTTATGCTAAAGTTAATAAAGAAAGCAAGCGCAAGCAAGCTAGCCTTGAAGA 480
Db 24581 GAACCAATTTATGCTAAAGTTAATAAAGAAAGCAAGCGCAAGCTAGCCTTGAAGA 24640
Qy 481 CCGATTAGCGCTCAAGTTGCTTAAAAAGTAAATGCAAAATTTGACCACTCAATCAATA 540
Db 24641 CCGATTAGCGCTCAAGTTGCTTAAAAAGTAAATGCAAAATTTGACCACTCAATCAATA 24700
Qy 541 GCAAGTGTGTTGGGTGTTGAGGCAAGCAACGGGCTTCCCTTGAAGAAAGGCTGATTA 600
Db 24701 GCAAGTGTGTTGGGTGTTGAGGCAAGCAACGGGCTTCCCTTGAAGAAAGGCTGATTA 24760
Qy 601 GTTGATGATCTCAAGTAAAGTAAAGGCTTTTCAAGCAATCAAGAA 642
Db 24761 GTTGATGATCTCAAGTAAAGTAAAGGCTTTTCAAGCAATCAAGAA 24802

RESULT 14
AB090088 3543 bp DNA linear BCT 07-NOV-2002
DEFINITION Helicobacter pylori gene for CagA, complete cds, strain:OK12.
ACCESSION AB090088
VERSION AB090088.1 GI:22335784
KEYWORDS
SOURCE Helicobacter pylori
ORGANISM Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteraceae; Helicobacter.
REFERENCE
1 Higashi, H., Tautsami, R., Fujita, A., Yamazaki, S., Aaka, M., Azuma, T.
and Hatakeyama, M.

TITLE
Biological activity of the Helicobacter pylori virulence factor
CagA is determined by variation in the tyrosine phosphorylation
sites
Proc. Natl. Acad. Sci. U.S.A. 99 (22), 14428-14433 (2002)
JOURNAL
PUBMED
12391297
2 (bases 1 to 3543)
Yamazaki, A., Yamazaki, S. and Azuma, T.
AUTHORS
Submitted (19-AUG-2002) Takeshi Azuma, Fukui Medical University,
Second Department of Internal Medicine; Matsuocho, Yoshida-gun,
Fukui 910-1193, Japan (E-mail: azuma@fmsr.fukui-med.ac.jp,
Tel:81-776-61-8351, Fax:81-776-61-8110)
JOURNAL
TITLE
Direct Submission
SUBMITTER
Location/Qualifiers
1. .3543
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="OK12"
/db_xref="taxon:210"
1. .3543
/country="Japan"
1. .3543
/gene="caga"
1. .3543
/gene="caga"
/product="caga"
/transl_table=1
/protein_id="BAC10434.1"
/db_xref="GI:22335785"

FEATURES
source
1. .3543
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="OK12"
/db_xref="taxon:210"
1. .3543
/country="Japan"
1. .3543
/gene="caga"
1. .3543
/gene="caga"
/product="caga"
/transl_table=1
/protein_id="BAC10434.1"
/db_xref="GI:22335785"
/translation="MTNRTIDQPTTAAANPDQFTNNIQAFLKDVNAVASFPDPQK
PIYDKNDNDNQAEPDI SOLRETSNKKATPKKQVPSPIKSNELINKNDIDV
ESSTSPKRGDDPYRIPTSWHSNDKNTNRSINRMHTTQPIPIPDDEKKEFL
KSAKSPGICITIGNOIRPDQKXWGFDSIKKEKKEKKEKPTGGAALDIFLSFTDK
VSGVDVIMQPELPHVDPDIASTHIGKLPESRDLDERENFSEKFLIDMELMD
VGHKQOQNNVATINVAHKNSSGLVLAGEKSHDIEPERVSLVANGNGRGAQDNAT
IANKVDPEPILQNNKNDNLSSEKKEKPEPTEIKQFQDQSPYLDALGMDRLAFAVSK
DPRGATITERNGKDLSTYLKQYKQKDKDLBEKRVTLQCNLKHQGVNFVSNFKY
TNAASGRKQGVNTGNSHLENGFSKVAVENLNNLAIIGVVRDLDEKLIAGKGS
ROBAKRLKQDLSSNKEVLGKALFNFAVAEAANGNDVETQKQDLDEKLIAGKGS
EKGAVNLSKSSGNKRNKMEASQASQKEIPALINKNANDAAITVAQNLIKGTKE
LEQKLENINNDLQDFSEKSPDFKNGKQKPSVAEETLFAKLGSVLCGIGNPMISKYE
NAAALNPFKNGKQKPSVAEETLFAKLGSVLCGIGNPMISKYE
DPSGIBQLADLNFSKQLAQAQKNEEDFNKQKSAIYQSVKGVNGLTVNGGSLRA
EATLTSNFSDDIKELIMAKLGNFNNNNNGLNKNEIYAKVKKKSGQASPEEPIVQ
YAKRVNAKIDRLNQLASGLGVQQAQAEFFPLKHDVDVDSLKVGSRVSPPIYATIDL
GAPFPLKHDKVDVDSLKVGSRVSPPIYATIDL
TKHENVNLWEGAKRVPAISAKLDIYNTNSTRINSIDIKNGAINKETGMLTKQKPE
WIKLVNKRTIAHNVGQSVPLSEYDKIGFNQKMKQVDSFSEKTKLNNAVQKVGSGFY
ILANAFSTGYCCLARENTEHGIKNVTKGQFS"

ORIGIN
Query Match
Best Local Similarity 85.3%; Score 584.4; DB 1; Length 3543;
Matches 606; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
Qy 1 AAAAATGCAAAAATAAGATTTCAGACAGTAAACGCAAGCAAAAGCGACCTTGAATAAT 60
Db 2248 AAAAATGCAAAAATAAGATTTCAGACAGTAAACGCAAGCAAAAGCGACCTTGAATAAT 2307
Qy 61 TCCGTTAAAGATGTATCATCATCAATCAAAAGTAAACGATTAAGTTATCTCAATCA 120
Db 2308 TCCATTAAGATGTATCATCATCAATCAAAAGTAAACGATTAAGTTATCTCAATCA 2367
Qy 121 GCGGTATCATGTGCTTAAGCAACGGGTGATTTCACTAGAGTAAAGCAACGGTACCGCAT 180
Db 2368 GCGGTATCATGTGCTTAAGCAACGGGTGATTTCACTAGAGTAAAGCAACGGTACCGCAT 2427
Qy 181 CTCMAAAATTTCTCAAAAGACCAATTTGGCCCAACAGCTCAAAAATGAAGTCTCAAT 240
Db 2428 CTCMAAAATTTCTCAAAAGACCAATTTGGCTCAAAAGCTCAAAAATGAAGTCTCAAT 2487
Qy 241 GCTAGAAAATAATCTGAATATATCAATCCGTTAAGATGTTGTAATGAAACCTTAAT 300

Db	2488	ACTGGAAAAAAATTCGCACTAATACCAATCCGTTAAGAAATGCTGTAAACGGAACCTACTG	254
Qy	301	GGTAATGGGTATCTCAGACGACGACCAACTCTTTCTTAAAACTTTTGACATCAAG	360
Db	2548	GGTAATGGGTATCTTAAGACAGACGACCAACTCTTTCTTAAAACTTTTGACATCAAG	2607
Qy	361	AAAGAGTTGAAGCAAACTTGGAATTTCAATTAACAATTAATGACTCAAAAC	420
Db	2608	AAAGAGTTGAAGCAAACTTGGAATTTCAATTAACAATTAATGACTCAAAAC	2667
Qy	421	GAACCATTTATGCTAAAGTTAATAAAAAAGAAAGGAGGCAAGCAGTAGCCTTGAAGA	480
Db	2668	GAACCATTTATGCTAAAGTTAATAAAAAAGAAATCAGAGCAAGCAGTACCTTGAAGA	2727
Qy	481	CCCATTTACGCTCAAGTGGCTCTAAAAAGGTAATGCAAAAAATTTGACCGACTCAATCAATA	540
Db	2728	CCCATTTACGCTCAAGTGGCTCTAAAAAGGTAATGCAAAAAATTTGACCGACTCAATCAATA	2787
Qy	541	GCAAGTGGTTGGGAGTTGTAGGGAGACAGCGGGCTTCCTTTGAAAAAGCATGATATA	600
Db	2788	GCAAGTGGTTGGGAGTTGTAGGGAGACAGCGGGCTTCCTTTGAAAAAGCATGATATA	2847
Qy	601	GTTGATGATCTCAGTAAGGTAGGGCTTTCAAGAAATCAAGA	642
Db	2848	GTTGATGATCTCAGTAAGGTAGGGCGATCGCTTACCGCTGA	2889
RESULT 15			
LOCUS	AY330636S2	26771 bp	DNA linear BCT 25-MAR-2004
DEFINITION	Helicobacter pylori strain Cas2 cag pathogenicity island section 2,		
ACCESSION	AY330637		
VERSION	AY330637.1	GI:37811777	
KEYWORDS			
SEGMENT			
SOURCE			
ORGANISM			
	2 of 2		
	Helicobacter pylori		
	Helicobacter pylori		
	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;		
	Helicobacteraceae; Helicobacter.		
	1 (bases 1 to 26771)		
	Blomstergren, A., Lundin, A., Nilsson, C., Engstrand, L. and		
	Lundberg, J.		
	Comparative analysis of the complete cag pathogenicity island		
	sequence in four Helicobacter pylori isolates		
	Gene 328, 85-93 (2004)		
	15019987		
	2 (bases 1 to 26771)		
	Blomstergren, A., Lundin, A., Nilsson, C., Engstrand, L. and		
	Lundberg, J.		
	Direct Submission		
	Submitted (26-JUN-2003) Biotechnology, Royal Institute of		
	Technology, Alba Nova University Centre, Roslagsskullbacken 21,		
	S-106 91 Stockholm, Sweden		
FEATURES			
source			
	1..26771		
	location/Qualifiers		
	/organism="Helicobacter pylori"		
	/mol_type="genomic DNA"		
	/strain="Cas2"		
	/db_xref="taxon:210"		
	<1..26014		
	/note="cag pathogenicity island"		
	complement(<1..2025)		
	/gene="cag7"		
	/locus_tag="HP0527"		
	/note="cag-Y; jhp0476; orf13/14; virB10-like protein"		
	/codon_start=1		
	/transl_table=11		
	/product="Cag7"		
	/protein_id="AA03862.1"		
	/db_xref="GI:37811787"		
	/translation="WNEENDLQETSSKAAQOHSPODLSENETTANHPEDLLKESSPH		
	YLDMPTEQTNFDDKSBREYQMDSGENIETSESNSGLADKLFKYARKYLVNKRPTT		
misc_feature			
CDS			

[illegible]

CDS

CDS

```

/locus_tag="HP0531"
6409..7059
/gene="cao11"
/locus_tag="HP0531"
/note="caq-U_jhp0480, orf18"
/codon_start=1
/translation_table=1
/product="Caq1"
/protein_id="AAK03866.1"
/db_xref="GI:37811791"
/translation="MDITTEHGSNP.LNAPSNSQNDLILTLSDIYEPKSLGQKPPF
HEALNASEELKLLLEIKLPOEKRYELLMOIQGKQFIMAAVYFSLGIGLEHHLL
GLCMGSLFVLAIPFPLRTSKNDLVESLTKRLGLQFYPAAGVAVLPFLPHTLS
IYELIYLGISINDKLVFLVNLNLCIFMGVLLYKVIKVIKNPINKGLPFSKQEQ"
7072..7914

```

gene

CDS

[illegible]

CDS

```

complement(8116,.8715)
//gene="Cag13"
//locus_tag="HP0534"
//note="cagS; jhp0482; orf20"
//codon_start=1
//transl_table=1
//product="Cag13"
//protein_id="AA03868.1"
//db_xref="GI:37811933"
//translation="MSNNRKLFSFIADSDKDKKELISLEQENLNTDEKKIDQI
KTMHPEPMQWTKGALDKYKLVKNTKQDRAVISGVDFKVKVYRLLESTEMLHAI
TNPMLPSKDRSLIGIFLPSPSPIMFQVNSITREDSQIEIYQTKPLPLKLPVMTDE
ENKFLANIKITBOQVNDYVDDLMEKLEDKYKKEKGVQNA"
complement(9153,.9533)

```

IDS

complement(9153..9533)
 /gene="cag14"
 /locus_tag="HP0535"
 /note="cagQ; jhp0483"
 /codon_start=1
 /transl_table=1
 /product="Cag14"
 /protein_id="FAA03869.1"
 /db_xref="rGI:37811794"
 /translation="MPTKTRIRDPKQETTOPPKIGLIMKTIASLLGGGTNLPFTGL
 SSDFLSNIIINPFLIMKGLINLEALCKKRNFLPMNDINKFAEVLKNGFDSIKNGSAL
 VGNFGSSKSDKTAANKMSQVRL"
 complement(9966..10310)
 /gene="cag15"
 /locus_tag="HP0536"
 complement(9966..10310)

```

/locus_tag="HP0536"
/note="cagP, hp0484"
/codon_start=1
/transl_table=11
/product="Cag15"

```

```

/protein_id="AA03870.1"
/db_xref="GI:37811795"
/translation="MKRISKLONFLQPKSFENKHLDRKSLYGLPNISSIVYGLI
ALFSGAGVGLVYPIILFPAIIRKSFERYITTYILLVSLSTISKTYLISHAFTMKL
IMFTQMWPL"
10745..11875
/gene="cag16"
/locus_tag="HP0537"
10745..11875
/gene="cag16"

Query Match      85.3%   Score 584.4   DB 1   Length 26771
Best Local Similarity 94.4%   Pred. No. 1.6e-99
Matches 606; Conservative 0; Mismatches 36; Total 0; Gaps 0
```

2

1	AAAAATGGCAAAATATGAGATTTCAGCAAGGTAAACGAACGAAAAAGCAGCTTGAAAT	60	
b	24154	AAAAATGGCAAAATATGAGATTTCAGCAAGGTAAACGAACGAAAAAGCAGCTTGAAAT	24213
1	61	TCCGTTAAAGTGTGATCATCATCAAAAGGTAAACGATTAATGTATATCTCAATCA	120
b	24214	TCCATTTAAAGTGTATCATCATCAAAAGATTAAGATTGAACATCTCATCA	24273
1	121	GGCGATTCACGTGGCTAAAGCAACGGGTATTCAGTAGGGTAGAAGCAACGCTTAGCCGAT	180
b	24274	GCTGTTCATACCTACTTAAGGCAACGGGATTCAGTGGGGTAGAACAAGCCGTAGCGGAT	24333
1	181	CTCAAAATTTCTCAAGAGCAATTGGCCCAACAGCTCAAAAAATAGAAAGTCCCAT	240
b	24334	CTCAAGAAATTTCAAAAGGCAATTGGCTCAACAGCTCAAAAAATAGAAAGTTCAT	24393
1	241	GCATGAAAAAATCGAATATATCATCTCGTTAGAAATGTTCTTAATGCAACCTTCTC	300

301 GGTAAATGGGTTATCTCAAGCAGAGCCACAACCTTTCTTAAAAAATTGGGACATCAAG 360
24454 GGTAAATGGGTTATCTTAAAGCAGAGCCACAACCTTTCTTAAAAAATTGGGACATCAAG 24513
361 AAAGAGTTGAATGCAAAACCTTGGAATTTCAATACATTAACATTAATGCACTCAAAAC 420
24514 AAAGAGTTGGAAGCAAAATTTGGGAATTTCAATACATTAATTAATGCACTCAAAAC 24573
421 GAACCCATTATTGCTAAAGTTATATAAAAAGAAAGCAGGCAGACAGCTAGCCTTGAGAA 480
24574 GAACCCATTATTGCTAAAGTTATATAAAAAGAAAGCAGGCAGACAGCTAGCCTTGAGAA 24633
481 CCCATTAGCCTCAAGTTGTAAAAAGTTAAATGCAAAATTTACCGATCACTAATTAATTA 540

```

24694 CCAATTAGCGCTCAAGTGTGTAAAAAGGTAAATGCAAAAATTGACCGACTCATATATATA 24695
541 GCAAGTGTTTGGGTGTGTTGAGGSCAAGCAGCGGCTTCCCTTTGAAAAGCATATATA 600
24694 GCAAGTGTTTGGGTGTGTTGAGGSCAAGCAGCGGCTTCCCTTTGAAAAGCATATATA 24753
601 GTTGATGATCTCAGTAGAGGTAGGCGCTTTCAAGAAATCAAGAA 642
24754 GTTGATGATCTCAGTAGAGGTAGGCGCTTTCAAGAAATCAAGAA 24795

```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2006, 20:26:07 ; Search time 39 Seconds
(without alignments)
562.498 Million cell updates/sec

Title: US-09-360-685c-25

Sequence: 1 KKGKNDPFSKYTAQKSDLEN.....LSRNOELAQKIDNINQAVSE 228

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1083	96.1	1215	2	B48281 cytotoxin-associat
2	959.5	85.1	1186	2	C64588 cag pathogenicity
3	896.5	79.5	1167	2	B71924 cag island protein
4	140	12.4	6713	2	B89921 hypothetical prote
5	127.5	11.3	774	2	JC2299 cell surface glyco
6	125	11.1	2481	2	D90011 FmeB protein (limp
7	124	11.0	504	2	D71615 hypothetical prote
8	124	11.0	1302	1	JC6009 surface-located me
9	121	10.7	993	2	E90598 membrane nucleas
10	121	10.7	1125	2	E90598 membrane nucleas
11	121	10.7	2401	2	T28676 rhoptry protein -
12	119.5	10.6	821	2	S67087 hypothetical prote
13	119	10.6	615	2	AB2641 methyl-accepting c
14	119	10.6	622	2	D97423 mclA protein (limp
15	118	10.5	1713	2	A55347 adhesive ligand ep
16	117.5	10.4	482	2	C86322 hypothetical prote
17	117	10.4	672	2	S61463 p83/100 protein -
18	117	10.4	693	2	S61464 p83/100 protein -
19	117	10.4	693	2	S61464 p83/100 protein -
20	116	10.3	667	2	B97012 p93 protein - Lym
21	116	10.3	1175	2	D35815 myosin heavy chain
22	116	10.3	1175	2	D35815 myosin heavy chain
23	116	10.3	1201	2	A35815 myosin heavy chain
24	116	10.3	1201	2	A35815 myosin heavy chain
25	116	10.3	2285	2	A32491 myosin heavy chain
26	116	10.3	2411	2	B32491 myosin heavy chain
27	115.5	10.2	1269	2	F84730 probable myosin he
28	115	10.2	1191	2	B97116 chromosome segrega
29	114.5	10.2	481	2	T18465 hypothetical prote

30	114.5	10.2	1051	2	T18351	limp protein - Myc
31	114.5	10.2	1120	2	A10561	integral membrane
32	114.5	10.2	1365	2	T30822	limp protein - Myc
33	114.5	10.2	1939	1	A67672	myosin alpha heavy
34	114	10.1	1005	2	A64465	hypothetical prote
35	114	10.1	3672	2	T23433	hypothetical prote
36	114	10.1	3704	2	T37316	probable laminin a
37	114	10.1	3712	2	S18253	laminin alpha-1 ch
38	113.5	10.1	722	2	H97217	uncharacterized co
39	113.5	10.1	955	1	A35254	leukotoxin A - Pan
40	113.5	10.1	2139	2	T18296	myosin heavy chain
41	113	10.0	1039	2	S62509	probable vesicular
42	113	10.0	1044	2	T50213	mechanosensitive C
43	113	10.0	1120	2	F90693	mechanosensitive C
44	113	10.0	1120	2	B85544	mechanosensitive C
45	113	10.0	1640	2	D86798	prophage p13 prote

ALIGNMENTS

RESULT 1

B48281

Cytotoxin-associated gene A protein cagA - Helicobacter pylori (strains G396 and CCUG 1

N/Alternative names: immunodominant 128k antigen; immunodominant 15k antigen

RESULT 2

C64588
 cag pathogenicity island protein cag26 - Helicobacter pylori (strain 26695)
 N:Alternate names: cytotoxin-associated gene A protein
 C:Species: Helicobacter pylori
 C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
 C/Accession: C64588
 R:Tomb, J.F.; White, O.; Kervavage, A.R.; Clayton, R.A.; Sutton, G.G.; Plieschmann, R.D.; Peterson, S.; Loftis, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodde, A.; McKenna, N.; J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; A. Authors: Mallin, B.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Frazer, C.
 A>Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 A:Reference number: A64520; M01D:97394467; PMID:9252185
 A/Accession: C64588
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 11186 <10M>
 A:Cross-references: UNIPROT:P55980; UNIPARC:UPI000126DD8; GB:AE000569; GB:AE000511; NID

Query Match
 Best Local Similarity 85.1%; Score 959.5; DB 2; Length 1186;
 Matches 205; Conservative 2; Mismatches 20; Indels 37; Gaps 3;

QY 1 KNGKNKDFSKVTQAKSDLENSVQVITINOKYTDKVDNLNOAVSVAKATGDSRYEQALAD 60
 DB 755 KNGKNKDFSKVTQAKSDLENSVQVITINOKYTDKVDNLNOAVSVAKATGDSRYEQALAD 814
 QY 61 LKNFSKEQLAQQAQKNEESLNARKKSEIYQSVKNGVNGTIVNGISQAAATTLSTNFSDIK 120
 DB 815 LKNFSKEQLAQQAQKNEEDPVTGKNSSELYQSVNGVNTIVNGISGIRATLAKNFSDIK 874
 QY 121 KEINAKLGNFNNNNNGNLKN--EPYAKVKKKKGAASLEPIYAVQAKVNAKIDRLN 178
 DB 875 KEINAKLGNF--NNNNNGLNKSTEPYAKVKKKKGAASLEPIYAVQAKVNAKIDRLN 933
 QY 179 QIASGLGVVGOAAGPFLKRDHDKVDLSKVGSLASPEPIYATIDLGSPFLKRDHDKVDL 209
 DB 934 QIASGLGVVGOAAGPFLKRDHDKVDLSKVGSLASPEPIYATIDLGSPFLKRDHDKVDL 209
 QY 210 -----SRNQLAOKIDNLNOAVSE 228
 DB 994 SKVGRSRNQLAOKIDNLNOAVSE 1017

RESULT 3

B71924
 cag island protein, cytotoxicity associated immunodominant antigen - Helicobacter pylori
 C:Species: Helicobacter pylori
 A:Variety: strain 999
 C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
 C/Accession: B71924
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Wang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999
 A>Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 A:Reference number: A71800; M01D:99120557; PMID:9923682
 A/Accession: B71924
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1167 <ARN>
 A:Cross-references: UNIPROT:Q9ZUT1; UNIPARC:UPI000126DD7; GB:AE001483; GB:AE001439; NID
 A:Experimental source: strain 999
 A:Gene: caga

Query Match
 Best Local Similarity 79.5%; Score 896.5; DB 2; Length 1167;
 Matches 189; Conservative 7; Mismatches 15; Indels 51; Gaps 2;

QY 1 KNGKNKDFSKVTQAKSDLENSVQVITINOKYTDKVDNLNOAVSVAKATGDSRYEQALAD 60
 DB 754 KNGKNKDFSKVTQAKSDLENSVQVITINOKYTDKVDNLNOAVSVAKATGDSRYEQALAD 813

QY 61 LKNFSKEQLAQQAQKNEESLNARKKSEIYQSVKNGVNGTIVNGISQAAATTLSTNFSDIK 120
 DB 814 LKNFSKEQLAQQAQKNEEDPVTGKNSSELYQSVNGVNTIVNGISGIRATLAKNFSDIK 873
 QY 121 KEINAKLGNFNNNNNGNLKNKEPIYAKVKKKKGAASLEPIYAVQAKVNAKIDRLN 180
 DB 874 KEINAKLGNFNNNNNGNLKNS--TEPIYAVQAKVNAKIDRLN 916
 QY 181 ASGLGVVGOAAG-----PFLKRDHDKVDLSK 206
 DB 917 ASGLGVVGOAAGPFLKRDHDKVDLSKVGSLASPEPIYATIDLGSPFLKRDHDKVDLSK 976
 QY 207 VGLSRNQLAOKIDNLNOAVSE 228
 DB 977 VGLSRNQLAOKIDNLNOAVSE 998

RESULT 4

B89921
 hypothetical protein ebha [imported] - Staphylococcus aureus (strain N315)
 C:Species: Staphylococcus aureus
 C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
 C/Accession: B89921
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Gai, L.; Ogu-
 ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 Lancet 357, 1225-1240, 2001
 A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A:Reference number: A89758; M01D:21311952; PMID:11418146
 A/Accession: B89921
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-6713 <10M>
 A:Cross-references: UNIPROT:Q99U54; UNIPARC:UPI000011021A; GB:BA000018; PID:g13701232; P
 A:Experimental source: strain N315
 A:Gene: ebha

Query Match
 Best Local Similarity 12.4%; Score 140; DB 2; Length 6713;
 Matches 66; Conservative 50; Mismatches 97; Indels 86; Gaps 11;

QY 5 NKPDSKVTQAKSD-----LENSVQVITINOKYTDKVDNLNOAVSVAKATGDSRY 54
 DB 5825 NKPDSKVTQAKSDLENSVQVITINOKYTDKVDNLNOAVSVAKATGDSRYEQALAD 5884
 QY 55 EQALADLKNFSK-----EQALAQKNEESLNARKK--SEIYQSVKNG 94
 DB 5885 AQLADLIDLVKAKEDAKQDVQKQVALIDEIDQPNLTDEKQALKDRINILDOGHNG 5944
 QY 95 VNGTLVNGLSQAAATTLSTNFSDIK-----KEINAKLGNFNNNN-- 135
 DB 5945 INNAATKEEIEOAKA-QLAQALKEIKDLYAKENAKQDVQKQVALIDEIDQPNLTDEK 6003
 QY 136 -----NGKNEPIYAKVKKKKGAASLEPIYAVQAKVNAKIDRLN 178
 DB 6004 KQALKDRINILDOGHNDINNAATKEEIEOAKAQLADLDO-----IDLVKAKEDAKNA 6058
 QY 179 -----QIASGLGVVGOAAGPFLKRDHDKVDLSKVGSLASPEPIYATIDLGSPFLKRDHDKVDLSK 227
 DB 6059 IKALANAKRDQINSNPDLTPEQVAKALK--EIDAEKRAL-QVENSNAQTIDQLNRGLN 6113

RESULT 5

JC2299
 cell surface glycoprotein MSG99 - Pneumocystis carinii
 C:Species: Pneumocystis carinii
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
 C/Accession: JC2299
 R:Wada, M.; Nakamura, Y.; DNA Res. 1, 163-168, 1994
 A>Title: MSG gene cluster encoding major cell surface glycoproteins of rat Pneumocystis

A:Reference number: JG2299; MUID:96051989; PMID:8535973
A:Accession: JG2299
A:Molecule type: DNA
A:Residues: 1-774 <MAD>
A:Cross-references: UNIPROT:Q01817; UNIPARC:UPI00000694BC; GB:D31909; GB:D17441; NID:955
C:Genetics:
A:Gene: MSG99
C:Superfamily: Pneumocystis carinii major surface glycoprotein MSG100
C:Keywords: glycoprotein

Query Match 11.3%; Score 127.5; DB 2; Length 774;
Best Local Similarity 22.4%; Pred. No. 12;
Matches 55; Conservative 46; Mismatches 95; Indels 49; Gaps 10;

3 GKNDPFSVTKQAKSDLENSVKDVIINQKVTDKVDNLNQAASVAKATGDF-----51
86 GIDKRYTSKVEICNNIEQKCD--IKAKIEKKIDIEGILITLITATKSIFYEENKYPEDTC 143
52 ----SRVEQALADLNKFSKEDL---AQQAQKNE-----SLNAR-KKSEIYGVKNGVGLT 99
144 MFPEPTKQAOYIDFCTIFPDTCYLAHKQEMENEFPLRLGADITKSKPOEKKQCEIPTL 203
100 VNGLSQAEATLTSKNFSDIKKELNAKGFNNNNNNNGLKNEPIYAKVKKKGAQAASLE 159
204 IKGGDDLASLLENISFEDFNNDIDRFSGSLDLDNKRKEETCYELG-----252
160 EPIYAQAQKVKYAKIDRLNQAASGLGVV--GQAAGF-PLKRDYVDLDSKVL-----S 210
253 --IYSPLEKRCBEKKSGLKKLCEGKGIITYSPPIKGFPNTER--EWTLLEKLGLENLYKEG 308
211 NQDEL 215
309 RNRGL 313

RESULT 6
D90011
FmTB protein [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: D90011
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogund
ma, A.; Mizutani-di, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
L: Lantec 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: D90011
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2481 <KUR>
A:Cross-references: UNIPROT:Q990R6; UNIPARC:UPI000000CAD87; GB:BA000018; PID:g13701961; F
A:Experimental source: strain N315
C:Genetics:
A:Gene: fmbB(mrp)

Query Match 11.1%; Score 125; DB 2; Length 2481;
Best Local Similarity 21.7%; Pred. No. 12;
Matches 51; Conservative 45; Mismatches 77; Indels 62; Gaps 8;

4 KKNDFSKVTKQAKSDLENSVKDVIINQKVTDKVDNLNQAASVAKATGDFSRV-EQALAD--60
1374 QNQAIDWVTGATTEERKNAKDLVLKAKEKAYQDITL-----AQTNDVDTQIKQDAVDIQ 1428
61 -----LKNSSKQQLAQQAQKESLNARKKSEIYGVKNGVGLVNGLSQAEATLTS 113
1429 GITADTTIKDVAKELATKAEQKALIAQ-----TADATTEE 1465
114 KNFSDIKKELNAKGFNNNNNNNGLKNEPIYAKVKKKGAQAASLEPIYAQAQKVKYAK 173
1466 K--EQANQAOYDAQLTQGNQ-----IENAGSIDVDYNTAK-DNAIDAIPIQASIDVKTNR 1518
174 IDRLNQAASGLGVVQAAGFPLKRDYVDLDSKVLGRNOELAQKIDNLNQAASVE 228

[illegible]

F:1154-1190/Domain: tetratricopeptide repeat homology <TT4>

Query Match 11.0%; Score 124; DB 1; Length 1302;

Best Local Similarity 21.0%; Pred. No. 6.3;

Matches 67; Conservative 59; Mismatches 87; Indels 106; Gaps 14;

QY 1 KNGKNDPFGVTOAKSDLENSVQV--ITNQKTDKVDNLNQAVAKATGDFSRVQAL 58
 DB 356 KSNLDNLSFABQSSLKDKESMSANDLNTLTLEVEIKLNKNGQKEA--KFNLEQQR 413
 QY 59 ADLKNSKEQL-----AQAQKNESLNAKKSEIY----- 88
 DB 414 KNIENFLTDEVKNNPNVATLVKDLTNAKDAKKSVT--NSNKSIDIINAEBALIQALDANK 472
 QY 89 -----QSVKNGVNGTLVNGLSQAETTLKSNFSD-----IKKEINATLGNFN- 131
 DB 473 AKQVDENAKSIKEQLNALI-----DKANTLLPQLNDNSEIYAKESLMEITNAK 525
 QY 132 --NNNNNG-----LKNPIYAKVKKKGAQAASLE----- 160
 DB 526 AVNQNDNASMOSAKSLDDKTKIQNQ--LTFPKDKAKFKELQTKQIDNPLTDVVK 583
 QY 161 --PIYAOVAKKV--NAKIDRLNQIASGLGVGAAGPPIKR-----HDKVDLSKVG 209
 DB 584 NNPNVATLVKDLTNAKDKKSVTKSNKSEIIANDEIQALDRAKAKAQQIDENAK-- 640
 QY 210 SRNOELAKQIDNLNQAVSE 228
 DB 641 SIKEQLSDSITNANQLNK 659

RESULT 9

C90072

hypothetical protein SA2436 [imported] - *Staphylococcus aureus* (strain N315)

C:Species: *Staphylococcus aureus*

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004

C:Accession: C90072

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc

ma, A.; Mizutani-Ji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.

A:Reference number: A69758; WUID:21311952; PMID:11418146

A:Accession: C90072

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-993 <KUR>

A:Cross-references: UNIPROT:Q99025; UNIPARC:UPI00000CAB7A; GB:BA000018; PID:913702601; F

A:Experimental source: strain N315

C:Genetics:

A:Gene: SA2436

Query Match 10.7%; Score 121; DB 2; Length 993;

Best Local Similarity 25.0%; Pred. No. 7;

Matches 66; Conservative 45; Mismatches 79; Indels 74; Gaps 16;

QY 12 TOAKS-----DLENSVDVITNQKTDKVDNLNQAVAKATGDFSRVQALDANKS 65
 DB 457 TOAKAQQKDSQALRNISYGIADPKSPRESLDNVSGLEFYTOYN--QQPITLLEIE 514
 QY 66 KEQ---LAQQAOK-----NESLNARKSRIYGVNGVNGTLVNGLSQAET-- 110
 DB 515 KNEVDLSKEIDKYAANNRINSL--RLVQLSWALNASSGT-----AAETGLAD 564
 QY 111 TLSK-----NPSD-IKKEINATLGNFN-----NNNNGLKNEPIYAKV--KXA 152
 DB 565 QLSKLDSSLSFQVYKDDNSSLVSIKRIIMDELNGQTALSN--VOSKNTITDQVINS 622
 QY 153 GOAASLEPIYAOVAKKVNAKIDRLNQIASGL--GVGQAAGPPLKHKDVKVDLSK 206
 DB 623 GQA-----IUKGKTRIDRLQTVLPSIEQOYISAVKNAQANFPKVSIVAAANF 672
 QY 207 VGLSRNQ--ELAQKIDNLNQAVSE 228

DB 673 V---RNDLPQBLQRLTNATVAVNK 693

RESULT 10

E90598

membrane nuclease, lipoprotein [imported] - *Mycoplasma pulmonis* (strain UAB CTIP)

C:Species: *Mycoplasma pulmonis*

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004

C:Accession: E90598

R:Chambaud, I.; Hallig, R.; Ferris, S.; Barde, V.; Samson, D.; Galisson, F.; Moszer, I.;

Nucleic Acids Res. 29, 2145-2153, 2001

A:Title: The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulm*

A:Reference number: A99512; WUID:21267165; PMID:11353084

A:Accession: E90598

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1125 <KUR>

A:Cross-references: UNIPROT:Q99PM9; UNIPARC:UPI00000D464B; GB:AL445566; PID:914090108;

A:Experimental source: strain UAB CTIP

C:Genetics:

A:Gene: MYPV 6930

Query Match 10.7%; Score 121; DB 2; Length 1125;

Best Local Similarity 24.1%; Pred. No. 8.1;

Matches 66; Conservative 41; Mismatches 107; Indels 60; Gaps 13;

QY 2 NGRKND-----FSKYTOAKSDLENSVQVITNQKTDKVDN-----NOAVSVAKAT 48
 DB 581 NGRNDSKQNTSNRQTKNDLSEKQNLTKKPSNSNSVETKNETONNNSSTKDE 640
 QY 49 GDSRVEQ--ALADLKFSKEQLAQAQKNESLNAR-----KSEIYQSVKNGV 95
 DB 641 IDTSACTQDSTNSNKKKEKTNQVSTKNTSNNSNSTKQRENSSTKEIKSESN-V 699
 QY 96 NGTLVNGLSQAETTLKSNFSDIKKEINATLGNFNNNNNGLKNEPIYAKVKK----- 150
 DB 700 NNS--NSTNQE-----ENIDNKKEEISKESVNSNSTVTOQET--PETESQNVYI 750
 QY 151 --KAGAASLEPIYAOVAKKVNAKIDRLNQIASGLGVGAAGPPIKR-----HDKVDLS 205
 DB 751 IGKPNQSLNQNAIDVSAKV--KIGYVN--INESVGSASAKFAVAKYIDHNKLDVG 807
 QY 206 KVGLSRNOELAKQIDNLN-----QAVSE 228
 DB 808 IGGLVHEETLTKEEEMNKLKDSQSDKMWQVISE 841

RESULT 11

T28676

rhoptry protein - *Plasmodium yoelii* (fragment)

C:Species: *Plasmodium yoelii*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T28676; A45521

R:Shiba, K.A.; Keen, J.K.; Ogum, S.A.; Holder, A.A.

Mol. Biochem. Parasitol. 76, 329-332, 1996

A:Title: Comparison of two members of a multigene family coding for high-molecular mass

A:Reference number: Z20507; WUID:97077455; PMID:8920022

A:Accession: T28676

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2401 <STN>

A:Cross-references: UNIPROT:Q26216; UNIPARC:UPI000017B647; EMBL:U36927; NID:91041784; PI

Mol. Biochem. Parasitol. 42, 241-246, 1990

A:Title: Identification of the gene for a *Plasmodium yoelii* rhoptry protein. Multiple co

A:Accession: A45521

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 2260-2401 <KEE>

A:Cross-references: UNIPARC:UPI000017B648; GB:M34281

QY 76 NESLNARKEEYOSKGVN-----GTLVNGLSQAETTLTKNPFSDIKKELNKG 129
Db 293 ERHNDKQELDGOIEPAGETIASGLGRSLKSLRTETTPAGRLDRLTDPNESLN 352
QY 130 FNNN-----NNGLAKNEPIYAKVKKKAGQAASLEPIYA--QVAKVNAKIDR 176
Db 353 LRDLGQIREFTLLIQNSGIEIKQSSVDLSKRTENQAASLEETAATAVEITATVRSSAR 412
QY 177 LNOIASGLGVGOAA---GPELKHDKYDDLSKY-GLSRN-DELAKIDNL 222
Db 413 ARANEAVVTKQASDSSGSVS--NAVDAKRIEASRKIEQIIEVIDDI 461

RESULT 15

A55347
adhesive ligand ep11gtrn, alpha-3 chain form A precursor - human
C:Species: Homo sapiens (man)
C:Date: 06-Feb-1995 #sequence_revisions 06-Feb-1995 #text_change 09-Jul-2004
C:Accession: A55347
R:Ryan, W.C.; Tizard, R.; Vandevanter, D.R.; Carter, W.G.
J Biol Chem. 269, 22779-22787, 1994
A:Title: Cloning of the lamA3 gene encoding the alpha3 chain of the adhesive ligand ep11gtrn.
A:Reference number: A55347; WUID:94357926; PMID:8077230
A:Accession: A55347
A:Status: P:1
A:Molecule type: primary
A:Residues: 15713 <R1A>
A:Cross-references: UNIPROT:Q16787; UNIPARC:UPI00000349C3; GB:L34155; NID:9551596; PID:9551596
C:Superfamily: laminin alpha-4 chain; laminin G repeat homology; laminin-type EGF-like
F:67-114/Domain: laminin-type EGF-like homology <LE2>
F:1192-1534/Domain: laminin G repeat homology <LE4>

Query Match 10 5% Score 118; DB 2; Length 1713;
Best Local Similarity 21.28; Pred. No. 20;
Matches 55; Conservative 47; Mismatches 107; Indels 50; Gaps 9;

QY 5 NKDFSKVTQAKSDLENSKDVIIINOKVTQVNDLNOAVSAKATGDFSR---VEQALAD 60
Db 300 NNNVNRATQSAKEIDVIRIKVIRNVHILKQISGDEGNVPSGDFSRMAEAQRMRE 359
QY 61 L--KNFSKEQLAQQAQKNEISLNARKKSEIYQVKNVNGTLVNGLSQAETTLK--N 115
Db 360 LNRNRFQGLREARADKRBSQLLNRIRTWOKTHQEN-----NGLANSIRDSLNEYAK 414
QY 116 PSDIKKELNAKLGNFNNNNNNGLKNEPIYAKVKK-----KAGQAASLEE 160
Db 415 LSDLRARLQEAALQAQKNGLNQENBERLGAIQOVKEINSLODFTKYITPADSLIQT 474
QY 161 PIYAQVAKVNAKIDRLNIOIASGLGVGOAAGPFLKHDKYDDLS---KGLSRN--- 212
Db 475 NIALQLMEKSKQKEYEL--ASINERQ-----ELSDKYRELRSAGKTSLVLEAKH 525
QY 213 ---DELAKIDNLNOAVS 227
Db 526 ARSLQELAKOLEIRIKNAS 544

Search completed: March 6, 2006, 20:30:33
Job time : 40 secs

Tue Mar 7 12:58:23 2006

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2006, 20:41:18 ; Search time 163 Seconds

(without alignments)
584,448 Million cell updates/sec

Title: US-09-360-685c-25

Perfect score: 1127
Sequence: 1 KNGKNDPSKYTOAKSDLEN.....LSRNGELAKINDINQAYSE 228

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgn2_6/prodata/1/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubppa/US08_PUBCOMB.pep.*
3: /cgn2_6/prodata/1/pubppa/US09_PUBCOMB.pep.*
4: /cgn2_6/prodata/1/pubppa/US10_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubppa/US10B_PUBCOMB.pep.*
6: /cgn2_6/prodata/1/pubppa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1127	100.0	1147	3	US-09-921-157-5 Sequence 5, Appl 1
2	1127	100.0	1338	4	US-09-402-100-4 Sequence 4, Appl 1
3	896.5	79.5	1167	4	US-10-335-977-8634 Sequence 8634, Ap
4	896.5	79.5	1178	4	US-10-335-977-8634 Sequence 8634, Ap
5	896.5	79.5	1183	4	US-10-335-977-8635 Sequence 8635, Ap
6	140	12.4	6713	4	US-10-282-122A-43811 Sequence 43811, A
7	131.5	11.7	1404	5	US-10-732-923-3304 Sequence 3304, A
8	131	11.6	3533	4	US-10-282-122A-70177 Sequence 70177, A
9	129	11.4	1992	5	US-10-470-048B-410 Sequence 410, Appl
10	129	11.4	10498	5	US-10-470-048B-410 Sequence 440, Appl
11	127	11.3	837	3	US-09-815-242-5883 Sequence 5883, Ap
12	127	11.3	875	3	US-09-815-242-13080 Sequence 13080, A
13	127	11.3	1361	4	US-10-369-493-3209 Sequence 3209, Ap
14	127	11.3	2434	3	US-09-815-242-5835 Sequence 5835, Ap
15	127	11.3	6281	3	US-09-815-242-12996 Sequence 12996, A
16	126	11.2	2368	3	US-09-815-242-5635 Sequence 5635, Ap
17	126	11.2	2368	3	US-09-815-242-12899 Sequence 12899, A
18	125.5	11.1	2437	3	US-09-815-242-5834 Sequence 5834, Ap
19	125	11.1	2478	3	US-09-815-242-5816 Sequence 5816, Ap
20	125	11.1	2478	3	US-09-815-242-12967 Sequence 12967, A
21	125	11.1	2478	3	US-10-470-048B-320 Sequence 220, Appl
22	125	11.1	2481	4	US-10-282-122A-43762 Sequence 43762, A
23	124	11.0	5795	3	US-09-815-242-12610 Sequence 12610, A
24	123	10.9	10203	4	US-10-661-809-23 Sequence 23, Appl
25	123	10.9	10203	4	US-10-724-972A-4098 Sequence 4098, Ap
26	121.5	10.8	6641	4	US-10-282-122A-70580 Sequence 70580, A
27	121	10.7	993	4	US-10-282-122A-43875 Sequence 43875, A

28	120	10.6	3158	3	US-09-815-242-12611 Sequence 12611, A
29	119	10.6	998	4	US-10-282-122A-70450 Sequence 70450, A
30	118	10.5	724	4	US-10-282-122A-71401 Sequence 71401, A
31	118	10.5	1441	5	US-10-732-923-3352 Sequence 3352, Ap
32	118	10.5	1693	4	US-10-603-725-4 Sequence 4, Appl 1
33	118	10.5	1693	4	US-10-603-725-8 Sequence 8, Appl 1
34	118	10.5	1713	4	US-10-171-311-113 Sequence 113, Appl
35	118	10.5	1713	4	US-10-372-683-10 Sequence 10, Appl
36	118	10.5	1713	4	US-10-603-725-6 Sequence 6, Appl 1
37	118	10.5	1724	4	US-10-603-725-2 Sequence 2, Appl 1
38	117.5	10.4	1047	4	US-10-282-122A-44353 Sequence 44353, A
39	117.5	10.4	2025	3	US-09-815-242-5703 Sequence 5703, Ap
40	117	10.4	693	4	US-10-369-100-68 Sequence 68, Appl
41	116	10.3	2067	6	US-11-097-143-40167 Sequence 40167, A
42	116	10.3	2539	5	US-10-831-070-6 Sequence 6, Appl 1
43	115.5	10.2	1090	4	US-10-282-122A-56217 Sequence 56217, A
44	115	10.2	873	3	US-09-952-267-13 Sequence 13, Appl
45	115	10.2	873	5	US-10-872-768-13 Sequence 13, Appl

ALIGNMENTS

```

RESULT 1
US-09-921-157-5
Sequence 5, Application US/09921157
GENERAL INFORMATION:
APPLICANT: Covacci, Antonio
APPLICANT: Bagnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappelli, Rino
TITLE OF INVENTION: Helicobacter Pylori Cytotoxin Proteins Useful For
FILE REFERENCE: CHIR0315
CURRENT APPLICATION NUMBER: US/09/921,157
PRIOR FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 08/466,662
PRIOR FILING DATE: 1995-06-06
PRIOR APPLICATION NUMBER: 08/256,848
PRIOR FILING DATE: 1994-10-21
PRIOR APPLICATION NUMBER: 09/360,934
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: 08/471,491
PRIOR FILING DATE: 1995-06-06
PRIOR APPLICATION NUMBER: PCT/EP93/00472
PRIOR FILING DATE: 1993-03-02
PRIOR APPLICATION NUMBER: PCT/EP93/00158
PRIOR FILING DATE: 1993-01-25
PRIOR APPLICATION NUMBER: FI 92 A 000052
PRIOR FILING DATE: 1992-03-02
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 1147
TYPE: PRT
ORGANISM: Helicobacter pylori
US-09-921-157-5

Query Match      100.0%; Score 1127; DB 3; Length 1147;
Best Local Similarity 100.0%; Pred. No. 9,7e-78;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KNGKNDPSKYTOAKSDLENQVYINQKTDVNDINQAVAKATGDSRYEQALAD 60
DB 750 KNGKNDPSKYTOAKSDLENQVYINQKTDVNDINQAVAKATGDSRYEQALAD 809
QY 61 LKNSKEQLOAQAKNSISINARKSEIYQSVKNGVNGTGVNGSOAEATLTKNFSDIK 120
DB 810 LKNSKEQLOAQAKNSISINARKSEIYQSVKNGVNGTGVNGSOAEATLTKNFSDIK 869
QY 121 KLNKAKIGNNNNNNGIKNEPIYAKVKKKAGQAAASLEBPIYAQVAKKVAKIRLNQI 180

```

Tue Mar 7 12:58:23 2006

us-09-360-685c-25.rapbm

Page 2

Db 870 KEINAKLGNFNNNNNGLKNEPIYAKVKKGAQAASLEPIYAQVAKVNAKIDRLNOI 929
QY 181 ASGLGVVGOAAGFPLKRHDVDDLSTKGLSNRQELAKIDINLQAVSE 228
Db 930 ASGLGVVGOAAGFPLKRHDVDDLSTKGLSNRQELAKIDINLQAVSE 977

RESULT 2
US-09-402-100-4
Sequence 4, Application US/09402100
Patent No. US20010019834A1
GENERAL INFORMATION:
APPLICANT: Daewoong Pharmaceutical Co, LTD
APPLICANT: Kim, Byung-O
APPLICANT: Shin, Sung-Seup
APPLICANT: Yu, Young-Hyo
APPLICANT: Park, Myung-Hwan
APPLICANT: Choi, Deok-Joon
APPLICANT: Jung, Hyung-Jin
TITLE OF INVENTION: Recombinant Microorganisms Expressing Antigenic Proteins of Helioth
FILE REFERENCE: 0136/06140
CURRENT APPLICATION NUMBER: US/09/402,100
EARLIER APPLICATION NUMBER: KR 97-11950
EARLIER FILING DATE: 1998-09-27
EARLIER APPLICATION NUMBER: KR 97-11951
EARLIER FILING DATE: 1997-03-31
NUMBER OF SEQ ID NOS: 48
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 1338
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Caga/CTXA2B Chimeric protein
US-09-402-100-4

Query Match
Best Local Similarity 100.0%; Score 1127; DB 3; Length 1338;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KNGKNDPSKVTQAKSDLENSVYDVIINOKVTDPKVDNLNOAVSAKATGDFSRVEQALAD 60
Db 750 KNGKNDPSKVTQAKSDLENSVYDVIINOKVTDPKVDNLNOAVSAKATGDFSRVEQALAD 809
QY 61 LKNSFSEQLAQQAQKVESINARKKSEIYOSVKNVGNGLVNGLSQAKETTLSTKNFSDIK 120
Db 810 LKNSFSEQLAQQAQKVESINARKKSEIYOSVKNVGNGLVNGLSQAKETTLSTKNFSDIK 869
QY 121 KEINAKLGNFNNNNNGLKNEPIYAKVKKGAQAASLEPIYAQVAKVNAKIDRLNOI 180
Db 870 KEINAKLGNFNNNNNGLKNEPIYAKVKKGAQAASLEPIYAQVAKVNAKIDRLNOI 929
QY 181 ASGLGVVGOAAGFPLKRHDVDDLSTKGLSNRQELAKIDINLQAVSE 228
Db 930 ASGLGVVGOAAGFPLKRHDVDDLSTKGLSNRQELAKIDINLQAVSE 977

RESULT 3
US-10-335-977-8633
Sequence 8633, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS: 10031
ADDRESS: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston

STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 8633:
SEQUENCE CHARACTERISTICS:
LENGTH: 1167 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc.feature
LOCATION: (B) LOCATION 1...1167
SEQUENCE DESCRIPTION: SEQ ID NO: 8633:
US-10-335-977-8633

Query Match
Best Local Similarity 79.5%; Score 996.5; DB 4; Length 1167;
Matches 189; Conservative 7; Mismatches 15; Indels 51; Gaps 2;
QY 1 KNGKNDPSKVTQAKSDLENSVYDVIINOKVTDPKVDNLNOAVSAKATGDFSRVEQALAD 60
Db 754 KNGKNDPSKVTQAKSDLENSVYDVIINOKVTDPKVDNLNOAVSAKATGDFSRVEQALAD 813
QY 61 LKNSFSEQLAQQAQKVESINARKKSEIYOSVKNVGNGLVNGLSQAKETTLSTKNFSDIK 120
Db 814 LKNSFSEQLAQQAQKVESINARKKSEIYOSVKNVGNGLVNGLSQAKETTLSTKNFSDIK 873
QY 121 KEINAKLGNFNNNNNGLKNEPIYAKVKKGAQAASLEPIYAQVAKVNAKIDRLNOI 180
Db 874 KEINAKLGNFNNNNNGLKNEPIYAKVKKGAQAASLEPIYAQVAKVNAKIDRLNOI 916
QY 181 ASGLGVVGOAAGFPLKRHDVDDLSTKGLSNRQELAKIDINLQAVSE 228
Db 917 ASGLGVVGOAAGFPLKRHDVDDLSTKGLSNRQELAKIDINLQAVSE 977

RESULT 4
US-10-335-977-8634
Sequence 8634, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS: 10031
ADDRESS: LAHIVE & COCKFIELD

STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 8634:
SEQUENCE CHARACTERISTICS:
LENGTH: 1178 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc.feature
LOCATION: (B) LOCATION 1...1178
SEQUENCE DESCRIPTION: SEQ ID NO: 8634:
US-10-335-977-8634

Query Match 79.5%; Score 896.5; DB 4; Length 1178;
Best Local Similarity 72.1%; Pred. No. 5e-60;
Matches 189; Conservative 7; Mismatches 15; Indels 51; Gaps 2;

QY 1 KNGKNDPSKVTQAKSDLENSVQDVIINQKVTDKVDNINQAVSVAKATGDFSRVEQALAD 60
DB 765 KNGKNDPSKVTQAKSDLENSVQDVIINQKVTDKVDNINQAVSVAKATGDFSRVEQALAD 824
QY 61 LKNFSKEQLAQQAQKNEISINARKSEIYOSVKNQVNGTLVNGLSQAERTLLSKNPSDIK 120
DB 825 LKNFSKEQLAQQAQKNEIDFNTGKNSALYOSVKNQVNGTLVNGLSQAERTLLSKNPSDIK 884
QY 121 KEINAKLGPNPNNNNNGLKNEPIYAKVNKKKAGQAASLEPIYAVAKKVNAKIDRLNOI 180
DB 885 KEINAKLGPNPNNNNNGLKNEPIYAKVNKKKAGQAASLEPIYAVAKKVNAKIDRLNOI 927
QY 181 ASGLGVVGOAG-----PFLKRDKVDLSK 206
DB 928 ASGLGVVGOAGSFLKRDKVDLSKVLGSANHEPIYATIDLGPFPLKRDKVDLSK 987
QY 207 VGLSRNOELAQKIDNINQAVSE 228
DB 988 VGLSRNOELAQKIDNINQAVSE 1009

RESULT 5
US-10-335-977-8635
Sequence 8635, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 8635:
SEQUENCE CHARACTERISTICS:
LENGTH: 1183 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc.feature
LOCATION: (B) LOCATION 1...1183.
SEQUENCE DESCRIPTION: SEQ ID NO: 8635:
US-10-335-977-8635

Query Match 79.5%; Score 896.5; DB 4; Length 1183;
Best Local Similarity 72.1%; Pred. No. 5e-60;
Matches 189; Conservative 7; Mismatches 15; Indels 51; Gaps 2;

QY 1 KNGKNDPSKVTQAKSDLENSVQDVIINQKVTDKVDNINQAVSVAKATGDFSRVEQALAD 60
DB 770 KNGKNDPSKVTQAKSDLENSVQDVIINQKVTDKVDNINQAVSVAKATGDFSRVEQALAD 829
QY 61 LKNFSKEQLAQQAQKNEISINARKSEIYOSVKNQVNGTLVNGLSQAERTLLSKNPSDIK 120
DB 830 LKNFSKEQLAQQAQKNEIDFNTGKNSALYOSVKNQVNGTLVNGLSQAERTLLSKNPSDIK 889
QY 121 KEINAKLGPNPNNNNNGLKNEPIYAKVNKKKAGQAASLEPIYAVAKKVNAKIDRLNOI 180
DB 890 KEINAKLGPNPNNNNNGLKNEPIYAKVNKKKAGQAASLEPIYAVAKKVNAKIDRLNOI 932
QY 181 ASGLGVVGOAG-----PFLKRDKVDLSK 206
DB 933 ASGLGVVGOAGSFLKRDKVDLSKVLGSANHEPIYATIDLGPFPLKRDKVDLSK 992
QY 207 VGLSRNOELAQKIDNINQAVSE 228
DB 993 VGLSRNOELAQKIDNINQAVSE 1014

RESULT 6
US-10-282-122A-43811
Sequence 43811, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl

Tue Mar 7 12:58:23 2006

us-09-360-685c-25.1rpbm

Page 4

```

/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Foreyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282.122A
/ PRIOR FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-22
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ REMAINING PRIOR APPLICATION DATA REMOVED - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 43811
/ LENGTH: 6713
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-10-282-122A-43811

```

```

Query Match 12.4%; Score 140; DB 4; Length 6713;
Best Local Similarity 22.1%; Pred. No. 0.54; Indels 86; Gaps 11;
Matches 66; Conservative 50; Mismatches 97;
QY 5 NNDPSKVTQAKSD-----LENSYKDV---INOKYTDKVDNINQAVSAKATGDFSRV 54
DB 5825 NNDVQVQVQALIDEIRNPNLTDKEKQALDKRINQLOQGHNDNNALTEKEIEQAKQL 5884
QY 55 EQALADLNKFSK-----EQLAQQAQKESLNARKK-----SEIYQSVNG 94
DB 5885 AQAQDIDKLVAKEDAKQDVQVQALIDEIDQNPNTDKEKQALDKRINQLOQGHNG 5944
QY 95 VNGTLVNGSLGSAEATLTSKNFSDIK-----KELNAKLGNFNNNNN----- 135
DB 5945 INNMTKEIEQAKA-QLAQAKKEIKQLVAKENAKQDVQVQVQALIDEIDQNPNTDKE 6003
QY 136 -----NGLKNEPIYAKVKKKGAQGAASLEBPYIAQVAKVAKIDRLN- 178
DB 6004 KQALKRINQIQQGHNDINNMTKEIEQAKAQLAQALD-----IKQLVAKEDAKNA 6058
QY 179 -----QIASGLGVGGAAGFPLKXHDVDDLSTKVLGSHNQLAKIDNINQAVS 227
DB 6059 IKALANAKRQDINSNPDLTPEQAKAKL---EIDBAEKRAL-QVNNAGTIDQNGRLN 6113

```

RESULT 7
US-10-732-923-3304
; Sequence 3304 Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D

```

/ TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
/ FILE REFERENCE: 38-15(927)6C
/ CURRENT APPLICATION NUMBER: US/10/732.923
/ CURRENT FILING DATE: 2003-12-10
/ PRIOR APPLICATION NUMBER: 10/310,154
/ PRIOR FILING DATE: 2002-12-04
/ NUMBER OF SEQ ID NOS: 24149
/ SEQ ID NO 3304
/ LENGTH: 1404
/ TYPE: PRT
/ ORGANISM: Mycoplasma hominis
US-10-732-923-3304

```

```

Query Match 11.7%; Score 131.5; DB 5; Length 1404;
Best Local Similarity 20.6%; Pred. No. 0.34; Indels 39; Gaps 9;
Matches 51; Conservative 60; Mismatches 97;
QY 1 KXGKNDPSKVTQAKSDLENSYKDV---ITNOKYTDKVDNINQAVSAKAT-GDFSRVQEA 57
DB 112 KXKKGHEBDQKIVQAMQEFKESQKALGDLINSDGQVDSNAGOSIDNTNNNSSTEQI 171
QY 58 LADLNKFSKQVLAQQAQKESLNARKKSEIYQSVNGVNGTIVGNGL---SQAETTLK 114
DB 172 IQALSKINEAKKELQSIINNARNQEK--EVEFEKKQQLNLTLSNEIDNSKKADETALIK 229
QY 115 NFSDI-----KELNAKLGNFNNNNNGLKNEPIYAKVKKKGAQGAASLEBPY 163
DB 230 NTNVAVDSDIKTETKTEIEKALIESLTNKLNE-FKQGEKAWAVASFSKSKQLDKLD 288
QY 164 AQVAKVNAKIDRLNQLASGLGVGGAAGFPLKXHDVDDLSTKVG--LSRQELAKIDN 221
DB 289 SEDCKVDSNE--SQVLT-----KTKIDNSISIDINKTKOIEKAIKS 331
QY 222 LNQAVSE 228
DB 332 LTNKIND 338

```

```

RESULT 8
US-10-282-122A-70177
; Sequence 70177 Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Foreyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282.122A
/ PRIOR FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625

```

PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/267,636
 PRIOR FILING DATE: 2001-02-09
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 Remaining prior application data removed - See file wrapper or PALM.
 NUMBER OF SEQ ID NOS: 78614
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 70177
 LENGTH: 3533
 TYPE: PRT
 ORGANISM: Staphylococcus aureus
 US-10-282-122A-70177

Query Match 11.6%; Score 131; DB 4; Length 3533;

Best Local Similarity 22.0%; Pred. No. 1.2;
Matches 66; Conservative 46; Mismatches 100; Indels 88; Gaps 11;

QY 5 NPDFSKVTOAKSD-----LENSVVDV---INQKVTDKVDNLQAVSAKATGDFSRV 54
 DB 2645 NNDVQKQVALDEIDRNPNLTDKEQALKDRINQILQOGHGINNMTKEIEQAKQL 2704
 QY 55 EQALADLNKFSK-----EQLAQQAQKNSLNARKK-----SEIYQSVKNG 94
 DB 2705 AQAQLODIKDLVAKEDAKQDVQKQVALDEIDQNPNTDKEQALKDRINQILQOGHND 2764
 QY 95 VNGTLVNGLSQA-----EATLSKSPSIIKELNAKGNFNNNNN--- 135
 DB 2765 INNMTKEAIEQAKERLAQALODIKDLVAKEDAKK--DIDKRVQALDEIDQNPNTDK 2822
 QY 136 -----NGLKNEPIYAKVKKKAGAAASLEPIYAQVAKVNAKIDRLN 178
 DB 2823 EKQALKDRINQILQOGHNDINNMTKEIEQAAQALQAD-----IKDLVAKEDAKN 2877
 QY 179 -----QIASGLGVVQAAGPFLRRHDKVDLSKVGSRNOELAQKIDNINQAVS 227
 DB 2878 AIKALANAKRDQINSNPDLPPEQAKALK--EIDAEKRAL-QVVENAQITIDQNRGIN 2933

RESULT 9
 US-10-470-048B-81
 Sequence 81, Application US/10470048B
 Publication No. US2005003744A1
 GENERAL INFORMATION:
 APPLICANT: MEINKE ET AL.
 TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
 TITLE OF INVENTION: ANTIGENS TO A SPECIFIC PATHOGEN
 FILE REFERENCE: SONN:035US
 CURRENT APPLICATION NUMBER: US/10/470,048B
 CURRENT FILING DATE: 2003-07-25
 NUMBER OF SEQ ID NOS: 603
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 81
 LENGTH: 1992
 TYPE: PRT
 ORGANISM: Staphylococcus aureus
 FEATURE:
 NAME/KEY: MOD RES
 LOCATION: (1222)..
 OTHER INFORMATION: X = anything
 US-10-470-048B-81

Query Match 11.4%; Score 129; DB 5; Length 1992;
 Best Local Similarity 24.3%; Pred. No. 0.81;
 Matches 65; Conservative 40; Mismatches 71; Indels 92; Gaps 13;

QY 6 KDFSKVTO-AKSDLENSVVDV---INQKVTDK-----VNDLNO----- 40
 DB 1171 KDLVAKEDAKQDVQKQVALDEIDQNPNTDKEQALKDRINQILQOGHNDINAMTK 1230
 QY 41 -AVSAKATGDFSRVEQALADLNKFSK-----EQLAQQAQKNSLNARKK-- 84

DB 1231 EAIQAK-----ERLAQALQDIKDLVAKEDAKNDIDKRVQALDEIDQNPNTDKEKA 1285
 QY 85 -----SEIYQSVKNGVNTLVNGLSQAEATLSKNPFSDIKELNAKGNFNNNNNGIK 139
 DB 1286 LKDRINQILQOGHNDINNMTKEIEQAKA-QLAQALQDIK----- 1325
 QY 140 NEPIYAKVKKKAGAAASLEPIYAQVAKVNAKIDRLNQIASGLGVVQAAGPFLKRD 199
 DB 1326 -DLVAKEDAKNAIKALA-----NAKRDQIN--SNPDLPPEQAKALK-- 1365
 QY 200 KYDDLSTKVGSRNOELAQKIDNINQAVS 227
 DB 1366 EIDAEKRAL-QVVENAQITIDQNRGIN 1392

RESULT 10
 US-10-470-048B-440
 Sequence 440, Application US/10470048B
 Publication No. US2005003744A1
 GENERAL INFORMATION:
 APPLICANT: MEINKE ET AL.
 TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
 TITLE OF INVENTION: ANTIGENS TO A SPECIFIC PATHOGEN
 FILE REFERENCE: SONN:035US
 CURRENT APPLICATION NUMBER: US/10/470,048B
 CURRENT FILING DATE: 2003-07-25
 NUMBER OF SEQ ID NOS: 603
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 440
 LENGTH: 10498
 TYPE: PRT
 ORGANISM: Staphylococcus aureus
 FEATURE:
 NAME/KEY: MOD RES
 LOCATION: (9728)..
 OTHER INFORMATION: X = anything
 US-10-470-048B-440

Query Match 11.4%; Score 129; DB 5; Length 10498;
 Best Local Similarity 24.3%; Pred. No. 6.6;
 Matches 65; Conservative 40; Mismatches 71; Indels 92; Gaps 13;

QY 6 KDFSKVTO-AKSDLENSVVDV---INQKVTDK-----VNDLNO----- 40
 DB 9677 KDLVAKEDAKQDVQKQVALDEIDQNPNTDKEQALKDRINQILQOGHNDINAMTK 9736
 QY 41 -AVSAKATGDFSRVEQALADLNKFSK-----EQLAQQAQKNSLNARKK-- 84
 DB 9737 EAIQAK-----ERLAQALQDIKDLVAKEDAKNDIDKRVQALDEIDQNPNTDKEQA 9791
 QY 85 -----SEIYQSVKNGVNTLVNGLSQAEATLSKNPFSDIKELNAKGNFNNNNNGIK 139
 DB 9792 LKDRINQILQOGHNDINNMTKEIEQAKA-QLAQALQDIK----- 9831
 QY 140 NEPIYAKVKKKAGAAASLEPIYAQVAKVNAKIDRLNQIASGLGVVQAAGPFLKRD 199
 DB 9832 -DLVAKEDAKNAIKALA-----NAKRDQIN--SNPDLPPEQAKALK-- 9871
 QY 200 KYDDLSTKVGSRNOELAQKIDNINQAVS 227
 DB 9872 EIDAEKRAL-QVVENAQITIDQNRGIN 9898

RESULT 11
 US-09-815-242-5883
 Sequence 5883, Application US/09815242
 Patent No. US20020061369A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari L.
 APPLICANT: Zykkind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.

Tue Mar 7 12:58:23 2006

us-09-360-685c-25.rapbm

Page 6

```

/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ TITLE OF INVENTION: Identification of Essential Genes in
/ FILE REFERENCE: ELITRA 011A
/ CURRENT APPLICATION NUMBER: US/09/815.242
/ PRIOR FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 5883
/ LENGTH: 837
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-09-815-242-5883

```

```

Query Match
Best Local Similarity 11.3%; Score 127; DB 3; Length 837;
Matches 69; Conservative 32; Mismatches 93; Indels 78; Gaps 12;

QY 3 GKNRDSKYTOAKSDLENSVQVYI---NOKVTD---KVDMLNOAVSVA----- 45
DB 470 GNEOKANNVDAMDKLQSIADNATTKONQVYTDASQKKDYNNAVTTAGIIDDQTSIP 529
QY 46 -----KATGDSRYEALADLNKF--SKEQLAQQAQKNSLNARKKEIYQSVKNGV 95
DB 530 TLDPTVYINQAQGVSTTKNALNGNENLEAAKQAQASQSLSDNLNNKQ-----QVTTDOI 585
QY 96 NG-----TLVGN---GLSQAEATTLTSKPSDIKKEELNKLGNFNNNN 134
DB 586 NGAHVTDEANQIKONQONNTAMGLKQAIADKATVFTDADQ--AKQAAYNTAV 642
QY 135 NNGELNEPIYAKVKKKQAQASLEPIYQVAKKYNKIDRLNQIASGLGVYQQAAGFP 194
DB 643 TNA---ENIISKANGNATQAE-----VEQAIKQVNAKQALNGNAN----- 681
QY 195 LKRHDKVDLSKVGSLSRNOELAQKIDMLNOAV 226
DB 682 -VOHAKDEATVTLINSSNDLNQAK-DALKQOV 711

```

```

RESULT 12
US-09-815-242-13080
/ Sequence 13080, Application US/09815242
/ Patent No. US20020061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl L.
/ APPLICANT: Zyskind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ TITLE OF INVENTION: Identification of Essential Genes in
/ FILE REFERENCE: ELITRA 011A
/ CURRENT APPLICATION NUMBER: US/09/815.242
/ CURRENT FILING DATE: 2001-03-21

```

```

/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13080
/ LENGTH: 875
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-09-815-242-13080

```

```

Query Match
Best Local Similarity 11.3%; Score 127; DB 3; Length 875;
Matches 69; Conservative 32; Mismatches 93; Indels 78; Gaps 12;

QY 3 GKNRDSKYTOAKSDLENSVQVYI---NOKVTD---KVDMLNOAVSVA----- 45
DB 507 GNEOKANNVDAMDKLQSIADNATTKONQVYTDASQKKDYNNAVTTAGIIDDQTSIP 566
QY 46 -----KATGDSRYEALADLNKF--SKEQLAQQAQKNSLNARKKEIYQSVKNGV 95
DB 567 TLDPTVYINQAQGVSTTKNALNGNENLEAAKQAQASQSLSDNLNNKQ-----QVTTDOI 622
QY 96 NG-----TLVGN---GLSQAEATTLTSKPSDIKKEELNKLGNFNNNN 134
DB 623 NGAHVTDEANQIKONQONNTAMGLKQAIADKATVFTDADQ--AKQAAYNTAV 679
QY 135 NNGELNEPIYAKVKKKQAQASLEPIYQVAKKYNKIDRLNQIASGLGVYQQAAGFP 194
DB 680 TNA---ENIISKANGNATQAE-----VEQAIKQVNAKQALNGNAN----- 718
QY 195 LKRHDKVDLSKVGSLSRNOELAQKIDMLNOAV 226
DB 719 -VOHAKDEATVTLINSSNDLNQAK-DALKQOV 748

```

```

RESULT 13
US-10-369-493-3209
/ Sequence 3209, Application US/10369493
/ Publication No. US20030233675A1
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianfeng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ FILE REFERENCE: 38-10(52052)B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ CURRENT FILING DATE: 2003-02-28
/ PRIOR APPLICATION NUMBER: US 60/360,039
/ PRIOR FILING DATE: 2002-02-21
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 3209
/ LENGTH: 1361
/ TYPE: PRT
/ ORGANISM: Neurospora crassa
US-10-369-493-3209

```

```

Query Match
Best Local Similarity 11.3%; Score 127; DB 4; Length 1361;
Matches 24.4%; Pred. No. 0.72;

```

Matches 76; Conservative 40; Mismatches 92; Indels 104; Gaps 14;

QY 6 KDFSXYTOAKSD-LENSVQV-----IINOK-----VTDKDNINQAV 42

DB 1003 KDLTKKOSQTEADALNDIDIKMKSTKEKLGQDAPAKETVLAERKEIIGGLKDSINRLNODI 1062

QY 43 SVAKATGDSR--VEQALADLK-----NFSKEQLOAOQAO-----74

DB 1063 STGNATLDKREIIDDLDKODIKRANSTIDTLKRDVADKQALILAKTKDVAARAEALAK 1122

QY 75 -----KNEBLNARKSEIYOSVNGVNGTLVNGLSQAEAT-----TLKKNPSDI 119

DB 1123 ARIASKNAL- AKTEEAKAFENQVOTLIDQAKGLNDVATKTTQLAQBRAISKLNKOI 1181

QY 120 -----KKEINAKLGNFN-----NNNNNGK--NEPIYAK--VAKKAGQASLE 159

DB 1182 FDLKTDVTKLQKSLSTKDNALTKQAGEISGRDLAKLREBELAKKALAKKTEEASISLE 1241

QY 160 EPIYAOVAKKVNKID--RLNQIASGLVGQAGFPLKRDHDKVDLSKVGLSRRQELA 216

DB 1242 KVKKLTDERATGAKDVTSTDTOLA-----ODKDAISKLEKDIKAKLNQELS 1287

QY 217 QKIDNINQAVSE 228

DB 1288 TKDASLTQKTGE 1299

RESULT 14

US-09-815-242-5835

/ Sequence 5835, Application US/09815242

/ Patent No. US20020061569A1

/ GENERAL INFORMATION:

/ APPLICANT: Haselbeck, Robert

/ APPLICANT: Ohlsen, Karl U.

/ APPLICANT: Zyskind, Judith W.

/ APPLICANT: Wall, Daniel

/ APPLICANT: Trawick, John D.

/ APPLICANT: Carr, Grant J.

/ APPLICANT: Yamamoto, Robert T.

/ APPLICANT: Xu, H. Howard

/ TITLE OF INVENTION: Identification of Essential Genes in

/ FILE REFERENCE: ELITRA.011A

/ CURRENT APPLICATION NUMBER: US/09/815,242

/ PRIOR FILING DATE: 2001-03-21

/ PRIOR APPLICATION NUMBER: 60/191,078

/ PRIOR FILING DATE: 2000-03-21

/ PRIOR APPLICATION NUMBER: 60/206,848

/ PRIOR FILING DATE: 2000-05-23

/ PRIOR APPLICATION NUMBER: 60/207,727

/ PRIOR FILING DATE: 2000-05-26

/ PRIOR APPLICATION NUMBER: 60/242,578

/ PRIOR FILING DATE: 2000-10-23

/ PRIOR APPLICATION NUMBER: 60/253,625

/ PRIOR FILING DATE: 2000-11-27

/ PRIOR APPLICATION NUMBER: 60/257,931

/ PRIOR FILING DATE: 2000-12-22

/ PRIOR APPLICATION NUMBER: 60/269,308

/ NUMBER OF SEQ ID NOS: 14110

/ SOFTWARE: FastSeq for Windows Version 4.0

/ SEQ ID NO 5835

/ LENGTH: 2434

/ TYPE: PRT

/ ORGANISM: Staphylococcus aureus

US-09-815-242-5835

Query Match 11.3%; Score 127; DB 3; Length 2434;

Best Local Similarity 25.4%; Pred. No. 1.5;

Matches 69; Conservative 32; Mismatches 93; Indels 78; Gaps 12;

QY 3 GKNDFSKVTOAKSDLENSVKDVI--NOKVTD--KVDNINQAVSVA-----45

DB 2130 GNEQKANNVDAMDKRQSIADNATTKQNTTDSQNKQAYNNAVTTAGIIDDQTSF 2189

QY 46 -----KATGDFSRVEQALADLNK--SKEQLOAOQKNEBLNARKKSEIYOSVNGV 95

DB 2190 TLDPFVINQAGVSTTKNALNGNENLEAKQAGQASQISGLSDLNNAOK--QVTTQDI 2245

QY 96 NG-----TLGVN---GLSQAEATTLKSNPSDIKELNKLGNFNNNN 134

DB 2246 NGAHTVDEANQIKQNAQNLNTAMGNLKAIDAKDATKATVNTFDADQ--AKQAVNTAV 2302

QY 135 NNGELNKEPIYAKVKKKAGQASLEBEPIYAOVAKKVNKIDRLNQIASGLVGQAGAPF 194

DB 2303 TNA--ENIISKANGNATQAE-----VEQALIKQVNAKQALNGNAN-----2341

QY 195 LKRHDKVDLSKVGLSRRQELAQKIDNINQAV 226

DB 2342 -VQHAKEBATLINSNDLNQAK--DALKQV 2371

RESULT 15

US-09-815-242-12996

/ Sequence 12996, Application US/09815242

/ Patent No. US20020061569A1

/ GENERAL INFORMATION:

/ APPLICANT: Haselbeck, Robert

/ APPLICANT: Ohlsen, Karl U.

/ APPLICANT: Zyskind, Judith W.

/ APPLICANT: Wall, Daniel

/ APPLICANT: Trawick, John D.

/ APPLICANT: Carr, Grant J.

/ APPLICANT: Yamamoto, Robert T.

/ APPLICANT: Xu, H. Howard

/ TITLE OF INVENTION: Identification of Essential Genes in

/ FILE REFERENCE: ELITRA.011A

/ CURRENT APPLICATION NUMBER: US/09/815,242

/ PRIOR FILING DATE: 2001-03-21

/ PRIOR APPLICATION NUMBER: 60/191,078

/ PRIOR FILING DATE: 2000-03-21

/ PRIOR APPLICATION NUMBER: 60/206,848

/ PRIOR FILING DATE: 2000-05-23

/ PRIOR APPLICATION NUMBER: 60/207,727

/ PRIOR FILING DATE: 2000-05-26

/ PRIOR APPLICATION NUMBER: 60/242,578

/ PRIOR FILING DATE: 2000-10-23

/ PRIOR APPLICATION NUMBER: 60/253,625

/ PRIOR FILING DATE: 2000-11-27

/ PRIOR APPLICATION NUMBER: 60/257,931

/ PRIOR FILING DATE: 2000-12-22

/ PRIOR APPLICATION NUMBER: 60/269,308

/ NUMBER OF SEQ ID NOS: 14110

/ SOFTWARE: FastSeq for Windows Version 4.0

/ SEQ ID NO 12996

/ LENGTH: 6281

/ TYPE: PRT

/ ORGANISM: Staphylococcus aureus

US-09-815-242-12996

Query Match 11.3%; Score 127; DB 3; Length 6281;

Best Local Similarity 25.4%; Pred. No. 4.9;

Matches 69; Conservative 32; Mismatches 93; Indels 78; Gaps 12;

QY 3 GKNDFSKVTOAKSDLENSVKDVI--NOKVTD--KVDNINQAVSVA-----45

DB 2909 GNEQKANNVDAMDKRQSIADNATTKQNTTDSQNKQAYNNAVTTAGIIDDQTSF 2968

QY 46 -----KATGDFSRVEQALADLNK--SKEQLOAOQKNEBLNARKKSEIYOSVNGV 95

DB 2969 TLDPFVINQAGVSTTKNALNGNENLEAKQAGQASQISGLSDLNNAOK--QVTTQDI 3024

QY 96 NG-----TLGVN---GLSQAEATTLKSNPSDIKELNKLGNFNNNN 134

Tue Mar 7 12:58:23 2006

us-09-360-685c-25.rapbm

Page 8

Db 3025 NGAHTVDEANQIKONQNLNTAMGNLKOALDADKDATKATVNFETDADQ--AKOQAYNTAV 3081
QY 135 NNGIKNEPIYAKVKKKKAGQASLEPIYAOVAKKVNAKIDRLNOIASGLGVVGQAAGFP 194
Db 3082 TNA--ENITISKANGNATQAE-----VEQALIKOVNAAKOALNGNAN----- 3120
QY 195 LKRHDKVDDLSKVGSLSPNOELAKTIDNLNQAV 226
Db 3121 -VOHAKDEFTALINSSWDLNOQAK-DALKQGV 3150

Search completed: March 6, 2006, 20:44:31
Job time : 164 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2006, 20:22:57 ; Search time 230 Seconds
(without alignments)
699.393 Million cell updates/sec

Title: US-09-360-685C-25

Perfect score: 1127

Sequence: 1 KNGKNDFSKVTQAKSDLEN.....LSRNOELAQKIDNINQAVSE 228

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: UniProt_05.80.*

1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1127	100.0	1147	1 CGA1_HELPY	P80200 heliobact
2	1127	100.0	1147	2 O9F220_HELPY	O9F220 heliobact
3	1035	91.8	1198	2 O6VRQ5_HELPY	O6VRQ5 heliobact
4	1029	91.3	1181	2 O6VRM7_HELPY	O6VRM7 heliobact
5	1029	91.3	1230	2 O4PL15_HELPY	O4PL15 heliobact
6	1025.5	91.0	267	2 O8RNU0_HELPY	O8RNU0 heliobact
7	1016	90.2	1142	2 O52GZ6_HELPY	O52GZ6 heliobact
8	1009	89.5	1180	2 O8KZH3_HELPY	O8KZH3 heliobact
9	1004	89.1	326	2 O7X4J1_HELPY	O7X4J1 heliobact
10	1003.5	89.0	1179	2 O9LSX9_HELPY	O9LSX9 heliobact
11	992	88.0	1186	2 O52GZ5_HELPY	O52GZ5 heliobact
12	990.5	87.9	1179	2 O9F222_HELPY	O9F222 heliobact
13	984	87.3	1182	2 O6VRG7_HELPY	O6VRG7 heliobact
14	982.5	87.2	394	2 O8RZB0_HELPY	O8RZB0 heliobact
15	972.5	86.3	327	2 O8RZB0_HELPY	O8RZB0 heliobact
16	972.5	86.3	1148	2 O60FP6_HELPY	O60FP6 heliobact
17	963	85.4	441	2 O9J7K3_HELPY	O9J7K3 heliobact
18	962.5	85.4	359	2 O8KZB8_HELPY	O8KZB8 heliobact
19	962.5	85.4	1183	2 O60FOS_HELPY	O60FOS heliobact
20	960	85.2	1183	2 O86064_HELPY	O86064 heliobact
21	959.5	85.1	1186	1 O86064_HELPY	O86064 heliobact
22	958.5	85.0	1147	1 O9F223_HELPY	O9F223 heliobact
23	958.5	85.0	1247	2 O9R115_HELPY	O9R115 heliobact
24	955.5	84.8	1156	2 O5D6E2_HELPY	O5D6E2 heliobact
25	952	84.5	1216	2 O8KZB9_HELPY	O8KZB9 heliobact
26	948.5	84.2	1247	2 O07910_HELPY	O07910 heliobact
27	944.5	83.8	408	2 O9J7K4_HELPY	O9J7K4 heliobact
28	942.5	83.6	1184	2 O60FQ2_HELPY	O60FQ2 heliobact
29	941	83.5	1222	2 O8KZB8_HELPY	O8KZB8 heliobact
30	939.5	83.4	359	2 O8KZB4_HELPY	O8KZB4 heliobact
31	939.5	83.4	1183	2 O60FP5_HELPY	O60FP5 heliobact

32	939.5	83.4	1183	2 O60FP7_HELPY	O60FP7 heliobact
33	939.5	83.4	1183	2 O60FQ3_HELPY	O60FQ3 heliobact
34	939.5	83.4	1188	2 O60FN3_HELPY	O60FN3 heliobact
35	938.5	83.3	1190	2 O5D6E1_HELPY	O5D6E1 heliobact
36	937.5	83.2	1183	2 O60FP9_HELPY	O60FP9 heliobact
37	934	82.9	391	2 O8KZC1_HELPY	O8KZC1 heliobact
38	934	82.9	1214	2 O60F08_HELPY	O60F08 heliobact
39	930.5	82.6	362	2 O8KZB5_HELPY	O8KZB5 heliobact
40	930.5	82.6	1183	2 O60FQ4_HELPY	O60FQ4 heliobact
41	930.5	82.6	1189	2 O60FP4_HELPY	O60FP4 heliobact
42	926.5	82.2	1320	2 O6VRJ6_HELPY	O6VRJ6 heliobact
43	918	81.5	1184	2 O60FP2_HELPY	O60FP2 heliobact
44	912.5	81.0	445	2 O9J7K5_HELPY	O9J7K5 heliobact
45	902.5	80.1	1182	1 CGA2_HELPY	P55746 heliobact

ALIGNMENTS

RESULT 1
CGA1_HELPY STANDARD; PRT; 1147 AA.
AC P80200;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Cytotoxicity associated immunodominant antigen (120 kDa protein).
GN Name=caga; Synonyms=cag;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteriales; Campylobacteriales;
OC Helicobacteriaceae; Helicobacter.
OX NCBI_Taxid=210;
NM (1)
NP NUCLEOTIDE SEQUENCE.
RC STRAIN=CCUG 17874 / NCTC 11638;
RX MEDLINE=93296225; PubMed=8516329;
RA Covacci A., Censini S., Bugnoli M., Petracca R., Burroni D.,
RA Macchia G., Massone A., Papini E., Xiang Z., Figura N., Rappuoli R.,
RT "Molecular characterization of the 128-kDa immunodominant antigen of
RT Helicobacter pylori associated with cytotoxicity and duodenal ulcer."
RT Proc. Natl. Acad. Sci. U.S.A. 90:5791-5795(1993).
RN (2)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CCUG 17874 / NCTC 11638;
RX MEDLINE=97121442; PubMed=8962108; DOI=10.1073/pnas.93.25.14648;
RA Censini S., Lange C., Xiang Z., Crabtree J., Chiara P., Borodovsky M.,
RA Rappuoli R., Covacci A.;
RT "cag, a pathogenicity island of Helicobacter pylori, encodes type I-
RT specific and disease-associated virulence factors."
RT Proc. Natl. Acad. Sci. U.S.A. 93:14648-14653(1996).
RN (3)
RP PROTEIN SEQUENCE OF 181-190; 314-328; 366-377; 420-430; 494-506;
RP 661-677; 900-914 AND 1062-1077.
RA Hermann V., Hermann J., Kist M.;
RT Submitted (APR-1993) to Swiss-Prot.
RT -!- FUNCTION: May be necessary for the transcription, folding, export,
RT or function of the cytotoxin.

This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.

EMBL: X70039. CAA49633.1; -; Genomic DNA.
EMBL: AF282853. AAC44706.1; -; Genomic DNA.
PIR: B48281; B48281.
InterPro: IPR005169; Caga.
InterPro: IPR004355; IVSec_caga.
PIfam: PF03507; Caga; 1.
DR PRINTS; PRO1553; TYPE4SSCAGA.
KW Antigen; Direct protein sequencing.
KW COMPBIA 880 885 Poly-Asn.

```
FT CONFLICT 320 320 G -> A (1n Ref. 3).
FT CONFLICT 325 325 P -> F (1n Ref. 3).
FT CONFLICT 328 328 R -> K (1n Ref. 3).
FT CONFLICT 426 426 K -> E (1n Ref. 3).
FT CONFLICT 429 429 O -> E (1n Ref. 3).
FT CONFLICT 673 675 AON -> TED (1n Ref. 3).
FT CONFLICT 901 901 A -> T (1n Ref. 3).
FT CONFLICT 903 903 L -> B (1n Ref. 3).
FT CONFLICT 907 907 L -> P (1n Ref. 3).
FT CONFLICT 910 910 P -> R (1n Ref. 3).
FT CONFLICT 914 914 P -> B (1n Ref. 3).
FT CONFLICT 1072 1072 P -> S (1n Ref. 3).
FT CONFLICT 1074 1074 S -> D (1n Ref. 3).
SQ SEQUENCE 1147 AA; 128014 MW; AB92770835F68490 CRC64;
```

Query Match 100.0%; Score 1127; DB 1; Length 1147;
Best Local Similarity 100.0%; Pred. No. 2, 1e-58;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 KNGKNDPSKYTOAKSDLENSVQVYINQKVTDKVDNLQAVSVAKATGDFSRVQALAD 60
DB 750 KNGKNDPSKYTOAKSDLENSVQVYINQKVTDKVDNLQAVSVAKATGDFSRVQALAD 809
QY 61 LKNFSKEQLAQQAQNESLNARKSEIYOSVKNVGVGLVNGLSQAETTLISKNFSDIK 120
DB 810 LKNFSKEQLAQQAQNESLNARKSEIYOSVKNVGVGLVNGLSQAETTLISKNFSDIK 869
QY 121 KEINAKLGPNNNNNGLKNEPIYAKVKKKAGQAASLEPIYAQVAKVNAKIDRLNOI 180
DB 870 KEINAKLGPNNNNNGLKNEPIYAKVKKKAGQAASLEPIYAQVAKVNAKIDRLNOI 929
QY 181 ASGLGVGOAAGPFLKRDHDKVDLSKVGLSRNQELAQKIDNLQAVSE 228
DB 930 ASGLGVGOAAGPFLKRDHDKVDLSKVGLSRNQELAQKIDNLQAVSE 977
```

RESULT 2
Q9F220 HELPY PRELIMINARY; PRT; 1147 AA.
ID Q9F220 HELPY PRELIMINARY; PRT; 1147 AA.
AC Q9F220 HELPY PRELIMINARY; PRT; 1147 AA.
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Caga.
GN Name=Caga;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
RN [1] TaxID=210;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCIC11637;
RX MEDLINE=21457648; PubMed=11573724; DOI=10.1007/s00550070002;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchiyama H.,
Ando T.;
RT "Heterogeneity found in the caga gene of Helicobacter pylori from
Japanese and non-Japanese isolates";
RL J. Gastroenterol. 35:890-897(2000).
DR GO; GO:0019534; F:toxin transposon activity; IEA.
DR InterPro; IPR005169; Caga.
DR InterPro; IPR004355; IVSec_caga.
DR Pfam; PF03507; Caga; 1.
DR PRINTS; PRO1553; TYPE4SCAGA.
SQ SEQUENCE 1147 AA; 128003 MW; D740N727549814ID CRC64;

Query Match 100.0%; Score 1127; DB 2; Length 1147;
Best Local Similarity 100.0%; Pred. No. 2, 1e-58;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 KNGKNDPSKYTOAKSDLENSVQVYINQKVTDKVDNLQAVSVAKATGDFSRVQALAD 60
DB 750 KNGKNDPSKYTOAKSDLENSVQVYINQKVTDKVDNLQAVSVAKATGDFSRVQALAD 809
```

```
QY 61 LKNFSKEQLAQQAQNESLNARKSEIYOSVKNVGVGLVNGLSQAETTLISKNFSDIK 120
DB 810 LKNFSKEQLAQQAQNESLNARKSEIYOSVKNVGVGLVNGLSQAETTLISKNFSDIK 869
QY 121 KEINAKLGPNNNNNGLKNEPIYAKVKKKAGQAASLEPIYAQVAKVNAKIDRLNOI 180
DB 870 KEINAKLGPNNNNNGLKNEPIYAKVKKKAGQAASLEPIYAQVAKVNAKIDRLNOI 929
QY 181 ASGLGVGOAAGPFLKRDHDKVDLSKVGLSRNQELAQKIDNLQAVSE 228
DB 930 ASGLGVGOAAGPFLKRDHDKVDLSKVGLSRNQELAQKIDNLQAVSE 977
```

RESULT 3
Q6VR05 HELPY PRELIMINARY; PRT; 1198 AA.
ID Q6VR05 HELPY PRELIMINARY; PRT; 1198 AA.
AC Q6VR05 HELPY PRELIMINARY; PRT; 1198 AA.
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Caga.
GN Name=Caga; ORFNames=HP0547;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
RN [1] TaxID=210;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C82;
RX PubMed=1501987; DOI=10.1016/j.gene.2003.11.029;
RA Blombergren A., Lundin A., Nilsson C., Engstrand L., Lundberg J.;
RT "Comparative analysis of the complete cag pathogenicity island
sequence in four Helicobacter pylori isolates";
RL Gene 328:85-93(2004).
DR EMBL; A130637; AAR03881.1; -; Genomic DNA.
DR GO; GO:0019534; F:toxin transposon activity; IEA.
DR InterPro; IPR005169; Caga.
DR InterPro; IPR004355; IVSec_caga.
DR Pfam; PF03507; Caga; 1.
DR PRINTS; PRO1553; TYPE4SCAGA.
SQ SEQUENCE 1198 AA; 133091 MW; DBBFSB554663BC0E CRC64;

Query Match 91.8%; Score 1035; DB 2; Length 1198;
Best Local Similarity 81.7%; Pred. No. 6, 2e-53;
Matches 214; Conservative 4; Mismatches 10; Indels 34; Gaps 1;

```
QY 1 KNGKNDPSKYTOAKSDLENSVQVYINQKVTDKVDNLQAVSVAKATGDFSRVQALAD 60
DB 766 KNGKNDPSKYTOAKSDLENSVQVYINQKVTDKVDNLQAVSVAKATGDFSRVQALAD 825
QY 61 LKNFSKEQLAQQAQNESLNARKSEIYOSVKNVGVGLVNGLSQAETTLISKNFSDIK 120
DB 826 LKNFSKEQLAQQAQNESLNARKSEIYOSVKNVGVGLVNGLSQAETTLISKNFSDIK 885
QY 121 KEINAKLGPNNNNNGLKNEPIYAKVKKKAGQAASLEPIYAQVAKVNAKIDRLNOI 180
DB 886 KEINAKLGPNNNNNGLKNEPIYAKVKKKAGQAASLEPIYAQVAKVNAKIDRLNOI 945
QY 181 ASGLGVGOAAGPFLKRDHDKVDLSKVGLSRNQELAQKIDNLQAVSE 228
DB 946 ASGLGVGOAAGPFLKRDHDKVDLSKVGLSRNQELAQKIDNLQAVSE 977
QY 207 VGLSRNQLAQKIDNLQAVSE 228
DB 1006 VGLSRNQLAQKIDNLQAVSE 1027
```

RESULT 4
Q6VRM7 HELPY PRELIMINARY; PRT; 1181 AA.
ID Q6VRM7 HELPY PRELIMINARY; PRT; 1181 AA.
AC Q6VRM7 HELPY PRELIMINARY; PRT; 1181 AA.
DT 05-JUL-2004 (TrEMBLrel. 27, Created)

DT 05-JUL-2004 (Tremblrel. 27, last sequence update)
 DT 05-JUL-2004 (Tremblrel. 27, last annotation update)
 DE Caga.
 GN Name=caga; ORFNames=HP0547;
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 NCBI_TaxId=210;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Ca73;
 RA PubMed=15019987; DOI=10.1016/j.gene.2003.11.029;
 RX Blomberg A., Lundin A., Nilsson C., Engstrand L., Lundberg J.;
 RT "Comparative analysis of the complete cag pathogenicity island
 RT sequence in four Helicobacter pylori isolates."
 RL Gene 328:85-93(2004).
 DR EMBL: AF330639; AAK03909.1; -; Genomic_DNA.
 DR GO: GO:0019534; P:toxin transporter activity; IBA.
 DR InterPro: IPR005169; Caga.
 DR Pfam: PF03507; Caga; 1.
 DR PRINTS: PR01553; TYPR4SSCAGA.
 DR SEQUENCE 1181 AA; 131427 MW; A991CAF6D17CDE3B CRC64;

Query Match 91.3%; Score 1029; DB 2; Length 1181;
 Best Local Similarity 81.3%; Pred. No. 1.4e-52;
 Matches 213; Conservative 4; Mismatches 11; Indels 34; Gaps 1;

QY 1 KNGKNDPSKVTQAKSDLENSVVDVITINQKVTDKVDNLNOAVSAKATGDFSRVBOALAD 60
 DB 750 KNGKNDPSKVTQAKSDLENSIKVDIINQKITDKVNLSAASVAKATGDFSRVBOALAD 809
 QY 61 LKNFSKEQLAQAQKKNESINARKSEIYGVNGVNGTLVNGLSQAERTTSSKPSDIK 120
 DB 810 LKNFSKEQLAQAQKKNESINARKSEIYGVNGVNGTLVNGLSQAERTTSSKPSDIK 869
 QY 121 KEINAKLGPNFNNNNNGLNKEPIYAKVKKKGAQAASLEPIYAQVAKKVNKIDRLNOI 180
 DB 870 KEINAKLGPNFNNNNNGLNKEPIYAKVKKKGAQAASLEPIYAQVAKKVNKIDRLNOI 929
 QY 181 ASGLGVVGOAAG-----FPLKRHDKVDLSK 206
 DB 930 ASGLGVVGOAAGFPLKRHDKVDLSKVGLSANPEPIYATIDELDGFPLKRHDKVDLSK 989
 QY 207 VGLSRNOELAKQKIDNLNOAVSE 228
 DB 990 VGLSRNOELAKQKIDNLNOAVSE 1011

RESULT 5

QAPLIS_HELPY
 ID QAPLIS_HELPY PRELIMINARY; PRT; 1230 AA.
 AC QAPLIS;
 DT 13-SEP-2005 (Tremblrel. 31, Created)
 DT 13-SEP-2005 (Tremblrel. 31, last sequence update)
 DT 13-SEP-2005 (Tremblrel. 31, last annotation update)
 DE Cytotoxin-associated protein A.
 GN Name=caga;
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 NCBI_TaxId=210;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=7d9e;
 RA Kim S.Y., Blaaser M.J., Lee Y.C., Pillinger M.H.;
 RT "Helicobacter pylori stimulates matrix metalloproteinase-1 secretion
 RT from gastric epithelial cells via Caga-dependent and independent
 RT mechanisms: requirement for Etk activation."
 RL Submitted (May-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL: DD067454; AAY68033.1; -; Genomic DNA.
 DR SEQUENCE 1230 AA; 136662 MW; 6AB240EB83001351 CRC64;

Query Match 91.3%; Score 1029; DB 2; Length 1230;
 Best Local Similarity 69.0%; Pred. No. 1.4e-52;
 Matches 218; Conservative 5; Mismatches 5; Indels 88; Gaps 1;

QY 1 KNGKNDPSKVTQAKSDLENSVVDVITINQKVTDKVDNLNOAVSAKATGDFSRVBOALAD 60
 DB 746 KNGKNDPSKVTQAKSDLENSIKVDIINQKITDKVNLSAASVAKATGDFSRVBOALAD 805
 QY 61 LKNFSKEQLAQAQKKNESINARKSEIYGVNGVNGTLVNGLSQAERTTSSKPSDIK 120
 DB 806 LKNFSKEQLAQAQKKNESINARKSEIYGVNGVNGTLVNGLSQAERTTSSKPSDIK 865
 QY 121 KEINAKLGPNFNNNNNGLNKEPIYAKVKKKGAQAASLEPIYAQVAKKVNKIDRLNOI 180
 DB 866 KEINAKLGPNFNNNNNGLNKEPIYAKVKKKGAQAASLEPIYAQVAKKVNKIDRLNOI 925
 QY 181 ASGLGVVGOAAG-----FPLKRHDKVDLSK 192
 DB 926 ASGLGVVGOAAGFPLKRHDKVDLSKVGSRVSEPIYATIDELDGFPLKRHDKVDLSK 985
 QY 193 -----FPLKRHDKVDLSK 212
 DB 986 VGOAAGFPLKRHDKVDLSKVGSRVSEPIYATIDELDGFPLKRHDKVDLSK 1045
 QY 213 OELAKIDNLNOAVSE 228
 DB 1046 OELAKIDNLNOAVSE 1061

RESULT 6

QBRND_HELPY
 ID QBRND_HELPY PRELIMINARY; PRT; 267 AA.
 AC QBRND;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, last annotation update)
 DE Cytotoxin-associated antigen (Fragment).
 GN Name=caga;
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 NCBI_TaxId=210;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=CH-CTX1;
 RA MEDLINE=20123532; PubMed=10660136;
 RX Opazo P., Muller I., Rollan A., Valenzuela P., Yudelevich A.,
 RA Garcia-de la Guardia R., Urra S., Venegas A.;
 RT "Serological response to Helicobacter pylori recombinant antigens in
 RT Chilean infected patients with duodenal ulcer, non-ulcer dyspepsia and
 RT gastric cancer."
 RL APMIS 107:1069-1078 (1999).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=CH-CTX1;
 RA MEDLINE=22121498; PubMed=12125208;
 RX Muller I., Medina-Selby A., Palacios J.L., Martinez P., Opazo P.,
 RA Bruce R., Mancilla M., Valenzuela P., Yudelevich A., Venegas A.;
 RT "Cloning and comparison of ten gene sequences of a Chilean H. pylori
 RT strain with other H. pylori strains revealed higher variability for
 RT vacA and caga virulence factors."
 RL Biol. Res. 35:67-84 (2002).
 DR EMBL: AF479032; AAL86902.1; -; Genomic_DNA.
 DR InterPro: IPR005169; Caga.
 DR Pfam: PF03507; Caga; 1.
 DR NON_TER 1
 FT NON_TER 267
 FT NON_TER 267
 SQ SEQUENCE 267 AA; 29032 MW; 04B4F51480960E8B CRC64;

Query Match 91.0%; Score 1025.5; DB 2; Length 267;
 Best Local Similarity 81.5%; Pred. No. 4.5e-53;
 Matches 216; Conservative 5; Mismatches 7; Indels 37; Gaps 3;

```

QY 1 KNGKNDPFSKYTOAKSDLENSYKDYIINOKITDKVDNLNOAVSVAKATGDFSRVEQALAD 60
DB 3 KNGKNDPFSKYTOAKSDLENSYKDYIINOKITDKVDNLNOAVSVAKATGDFSRVEQALAD 62
QY 61 LKNFSKEQLAQOAOQKNESLNARKKSEIYGVKNGVNGTLVNGLSQAEATLTLTKNFSDIK 120
DB 63 LKNFSKEQLAQOAOQKNESLNARKKSEIYGVKNGVNGTLVNGLSQAEATLTLTKNFSDIK 122
QY 121 KEINAKLGNFNPNNNNNNGLNKNEPIYAKVKKKGAQAASLEPIYAOVAKKNAKIDRL 177
DB 123 KEINAKLGNFNPNNNNNNGLNKNEPIYAKVKKKGAQAASLEPIYAOVAKKNAKIDRL 182
QY 178 NOIASGLGVGOAAG-----PFLKRHDKYDD 203
DB 183 NOIASGLGVGOAAGFPLKRHDKYDDLSKVGSRVSPPEPIYATIDDLGGFPLKRHDKYDD 242
QY 204 LSKVGLSHNOELAQKIDNLNOAVSE 228
DB 243 LSKVGLSHNOELAQKIDNLNOAVSE 267

```

```

RESULT 7
ID 052GZ6_HELPY PRELIMINARY; PRT; 1142 AA.
AC 052GZ6;
DT 13-SEP-2005 (Tremblrel. 31, Created)
DT 13-SEP-2005 (Tremblrel. 31, Last sequence update)
DE Caga (Fragment)
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OX NCBI_TaxId=210;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2808.
RA PubMed=15972330; DOI=10.1073/pnas.0409873102;
RT "NF- $\kappa$ B activation and potentiation of proinflammatory responses
RL Proc. Natl. Acad. Sci. USA. 102:9300-9305 (2005).
DR EMBL; D0011619; AAY18597.1; Genomic DNA.
DR GO; GO:0019534; F100xin transpporter activity; IEA.
DR InterPro; IPR005169; Caga.
DR Pfam; PF03507; Caga; 1.
DR PRINTS; PRO1553; TYPE4SSCAGA.
FT NON TER 1142
SQ SEQUENCE 1142 AA; 128233 MW; E53DBE9756EA29BE CRC64;

```

```

Query Match 90.2%; Score 1016; DB 2; Length 1142;
Best Local Similarity 89.0%; Pred No. 7.8e-52;
Matches 203; Conservative 7; Mismatches 18; Indels 0; Gaps 0;

```

```

QY 1 KNGKNDPFSKYTOAKSDLENSYKDYIINOKITDKVDNLNOAVSVAKATGDFSRVEQALAD 60
DB 745 KNGKNDPFSKYTOAKSDLENSYKDYIINOKITDKVDNLNOAVSVAKATGDFSRVEQALAD 804
QY 61 LKNFSKEQLAQOAOQKNESLNARKKSEIYGVKNGVNGTLVNGLSQAEATLTLTKNFSDIK 120
DB 805 LKNFSKEQLAQOAOQKNEDPNTGKNSLYGVKNGVNGTLVNGLSQAEATLTLTKNFSDIK 864
QY 121 KEINAKLGNFNPNNNNNNGLNKNEPIYAKVKKKGAQAASLEPIYAOVAKKNAKIDRLNOI 180
DB 865 KEINAKLGNFNPNNNNNNGLNKNEPIYAKVKKKGAQAASLEPIYAOVAKKNAKIDRLNOI 924
QY 181 ASGLGVGOAAGFPLKRHDKYDDLSKVGSRVSPPEPIYATIDDLGGFPLKRHDKYDDLSK 228
DB 925 ASGLGVGOAAGFPLKRHDKYDDLSKVGSRVSPPEPIYATIDDLGGFPLKRHDKYDDLSK 972

```

RESULT 8
08KZK3_HELPY

```

ID 08KZK3_HELPY PRELIMINARY; PRT; 1180 AA.
AC 08KZK3;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DE Caga (Cag pathogenicity island protein).
GN Name=caga; Synonyms=HP0547;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OX NCBI_TaxId=210;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OK112;
RA MEDLINE=2225025; PubMed=12391297; DOI=10.1073/pnas.222375399;
RA Higaehi H.; Teutsumi R.; Fujita A.; Yamazaki S.; Asaka M.; Azuma T.;
RA Harakeyama W.
RT "Biological activity of the Helicobacter pylori virulence factor Caga
is determined by variation in the tyrosine phosphorylation sites."
RL Proc. Natl. Acad. Sci. U.S.A. 99:14428-14433 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OK112.
RA Yamazaki S.; Azuma T.;
RA Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
RP STRAIN=OK112;
RP NUCLEOTIDE SEQUENCE.
RA Azuma T.; Yamazaki S.;
RA Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB090088; BAC10434.1; Genomic DNA.
DR GO; GO:0019534; F100xin transpporter activity; IEA.
DR InterPro; IPR005169; Caga.
DR Pfam; PF03507; Caga; 1.
DR PRINTS; PRO1553; TYPE4SSCAGA.
SQ SEQUENCE 1180 AA; 130973 MW; 62347A4F2635020 CRC64;

```

```

Query Match 89.5%; Score 1009; DB 2; Length 1180;
Best Local Similarity 79.0%; Pred No. 2.1e-51;
Matches 207; Conservative 9; Mismatches 12; Indels 34; Gaps 1;

```

```

QY 1 KNGKNDPFSKYTOAKSDLENSYKDYIINOKITDKVDNLNOAVSVAKATGDFSRVEQALAD 60
DB 750 KNGKNDPFSKYTOAKSDLENSYKDYIINOKITDKVDNLNOAVSVAKATGDFSRVEQALAD 809
QY 61 LKNFSKEQLAQOAOQKNESLNARKKSEIYGVKNGVNGTLVNGLSQAEATLTLTKNFSDIK 120
DB 810 LKNFSKEQLAQOAOQKNEDPNTGKNSLYGVKNGVNGTLVNGLSQAEATLTLTKNFSDIK 869
QY 121 KEINAKLGNFNPNNNNNNGLNKNEPIYAKVKKKGAQAASLEPIYAOVAKKNAKIDRLNOI 180
DB 870 KEINAKLGNFNPNNNNNNGLNKNEPIYAKVKKKGAQAASLEPIYAOVAKKNAKIDRLNOI 929
QY 181 ASGLGVGOAAG-----PFLKRHDKYDDLSK 206
DB 930 ASGLGVGOAAGFPLKRHDKYDDLSKVGSRVSPPEPIYATIDDLGGFPLKRHDKYDDLSK 969
QY 207 VELSHNOELAQKIDNLNOAVSE 228
DB 990 VELSHNOELAQKIDNLNOAVSK 1011

```

```

RESULT 9
ID 07XKJ1_HELPY PRELIMINARY; PRT; 326 AA.
AC 07XKJ1;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DE Caga (Fragment).
GN Name=caga;

```

Tue Mar 7 12:58:24 2006

OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=China:079a;
 RX PubMed=14734191; DOI=10.1016/S0928-8244(03)00299-2;
 RA Zhou W., Yamazaki S., Yamakawa A., Ohnishi M., Ito Y., Keida Y.,
 Higaishi H., Hatakeyama M., Si J., Azuma T., Ico Y.,
 RT "The diversity of vacA and cagA genes of Helicobacter pylori in East
 Asia".
 RL FEMS Immunol. Med. Microbiol. 40:81-87 (2004).
 DR EMBL; AB110963; BAC77006.1; -; Genomic_DNA.
 DR InterPro; IPR005169; Caga.
 DR Pfam; PF03507; Caga; 1.
 FT NON_TER 1 326
 FT NON_TER 326 326
 SQ SEQUENCE 326 AA; 35988 MW; 7BC601BFA98ACABF CRC64;

Query Match 89.1%; Score 1004; DB 2; Length 326;
 Best Local Similarity 88.2%; Pred. No. 1e-51;
 Matches 201; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 KNGKNDPSKYTQAKSDLENSVDVITINOKYTDVNDLNOAVSAKATGDFSRVEQALAD 60
 DB 57 KNGKNDPSKYTQAKSDLENSVDVITINOKYTDVNDLNOAVSAKATGDFSRVEQALAD 116
 QY 61 LKNFSKEQLAQOAKNESLNARKKSEIYOSVKNVNGTLVNGLSQAETTLTKNFSIDIK 120
 DB 117 LKNFSKEQLAQOAKNESLNARKKSEIYOSVKNVNGTLVNGLSQAETTLTKNFSIDIK 176
 QY 121 KEINAKLGNFNNNNNGKNEPIYAKVKKAGQAASLEPIYAOVAKKNAKIDRLNOI 180
 DB 177 KEINAKLGNFNNNNNGKNEPIYAKVKKAGQAASLEPIYAOVAKKNAKIDRLNOI 236
 QY 181 ASGLGVGQAGFPLKRHDKVDLSKYGSRNOELAQKIDNLNOAVSE 228
 DB 237 ASGLGVGQAGFPLKRHDKVDLSKYGSRNOELAQKIDNLNOAVSE 284

RESULT 10
 ID Q9LSX9_HELPY PRELIMINARY; PRT; 1179 AA.
 AC Q9LSX9_HELPY
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Cytotoxin-associated protein CagA.
 GN Name=cagA;
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=SS1;
 RA Zhang J.Z.;
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=SS1;
 RA Zhou J.C., Zhang J.Z.;
 RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF247651; AAF63759.1; -; Genomic DNA.
 DR GO; GO:0019534; F:toxin transporter activity; IEA.
 DR InterPro; IPR005169; Caga.
 DR InterPro; IPR004355; IVSec_caga.
 DR Pfam; PF03507; Caga; 2.
 DR PRINTS; PRO1553; TYPE4SSCAGA.
 SQ SEQUENCE 1179 AA; 131297 MW; 7A9112F58B749787 CRC64;

Query Match 89.0%; Score 1003.5; DB 2; Length 1179;

Best Local Similarity 79.8%; Pred. No. 4.4e-51;
 Matches 209; Conservative 5; Mismatches 13; Indels 35; Gaps 2;
 QY 1 KNGKNDPSKYTQAKSDLENSVDVITINOKYTDVNDLNOAVSAKATGDFSRVEQALAD 60
 DB 750 KNGKNDPSKYTQAKSDLENSVDVITINOKYTDVNDLNOAVSAKATGDFSRVEQALAD 809
 QY 61 LKNFSKEQLAQOAKNESLNARKKSEIYOSVKNVNGTLVNGLSQAETTLTKNFSIDIK 120
 DB 810 LKNFSKEQLAQOAKNESLNARKKSEIYOSVKNVNGTLVNGLSQAETTLTKNFSIDIK 869
 QY 121 KEINAKLGNFNNNNNGKNEPIYAKVKKAGQAASLEPIYAOVAKKNAKIDRLNOI 180
 DB 870 KEINAKLGNFNNNNNGKNEPIYAKVKKAGQAASLEPIYAOVAKKNAKIDRLNOI 929
 QY 181 ASGLGVGQAGFPLKRHDKVDLSKYGSRNOELAQKIDNLNOAVSE 228
 DB 930 ASGLGVGQAGFPLKRHDKVDLSKYGSRNOELAQKIDNLNOAVSE 988
 QY 207 VGLSRNOELAQKIDNLNOAVSE 228
 DB 989 VGLSRNOELAQKIDNLNOAVSE 1010

RESULT 11
 ID Q52GZ5_HELPY PRELIMINARY; PRT; 1186 AA.
 AC Q52GZ5_HELPY
 DT 13-SEP-2005 (Tremblrel. 31, Created)
 DT 13-SEP-2005 (Tremblrel. 31, Last sequence update)
 DT 13-SEP-2005 (Tremblrel. 31, Last annotation update)
 DE Caga (Fragment).
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=P310;
 RX PubMed=15972330; DOI=10.1073/pnae.0409873102;
 RA Brandt S., Kwok T., Hartig R., Konig W., Backett S.;
 RT "NF- κ B activation and potentiation of proinflammatory responses
 by the Helicobacter pylori CagA protein".
 RL Proc. Natl. Acad. Sci. U.S.A. 102:9300-9305 (2005).
 DR EMBL; DQ011620; AAY18598.1; -; Genomic DNA.
 DR GO; GO:0019534; F:toxin transporter activity; IEA.
 DR InterPro; IPR005169; Caga.
 DR InterPro; IPR004355; IVSec_caga.
 DR Pfam; PF03507; Caga; 1.
 DR PRINTS; PRO1553; TYPE4SSCAGA.
 FT NON_TER 1186 1186
 SQ SEQUENCE 1186 AA; 131865 MW; CD5AC9D1DA1BF3F CRC64;

Query Match 88.0%; Score 992; DB 2; Length 1186;
 Best Local Similarity 78.8%; Pred. No. 2.1e-50;
 Matches 212; Conservative 5; Mismatches 10; Indels 42; Gaps 3;

QY 1 KNGKNDPSKYTQAKSDLENSVDVITINOKYTDVNDLNOAVSAKATGDFSRVEQALAD 60
 DB 750 KNGKNDPSKYTQAKSDLENSVDVITINOKYTDVNDLNOAVSAKATGDFSRVEQALAD 809
 QY 61 LKNFSKEQLAQOAKNESLNARKKSEIYOSVKNVNGTLVNGLSQAETTLTKNFSIDIK 120
 DB 810 LKNFSKEQLAQOAKNESLNARKKSEIYOSVKNVNGTLVNGLSQAETTLTKNFSIDIK 869
 QY 121 KEINAKLGNFNNNNNGKNEPIYAKVKKAGQAASLEPIYAOVAKKNAKIDRLNOI 180
 DB 870 KEINAKLGNFNNNNNGKNEPIYAKVKKAGQAASLEPIYAOVAKKNAKIDRLNOI 929
 QY 174 IDRLNOIASGLGVGQAGFPLKRHDKVDLSKYGSRNOELAQKIDNLNOAVSE 228
 DB 929 IDRLNOIASGLGVGQAGFPLKRHDKVDLSKYGSRNOELAQKIDNLNOAVSE 988

Tue Mar 7 12:58:24 2006

us-09-360-685c-25.rup

Page 6

QY 200 KVDLSKYGLSRNOELAKIDINQAVSE 228
DB 989 KVDLSKYGLSRNOELAKIDINQAVSE 1017

RESULT 12

Q9F222 HELPY PRELIMINARY; PRT: 1179 AA.
AC Q9F222;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
GN Name=caga;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OX NCBI_Taxid=210;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=ATCC43579;
RX MEDLINE=21457648; PubMed=11573724; DOI=10.1007/s005350070002;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
Ando T.;
RT "Heterogeneity found in the caga gene of Helicobacter pylori from
Japanese and non-Japanese isolates.";
RL EMBL: AB015414; BAE20924.1; -; Genomic DNA.
DR GO: GO:0019534; F:toxin transporter activity; IEA.
DR InterPro: IPR004355; IYSec_caga.
DR Pfam: PF03507; Caga_1.
DR PRINTS: PRO1553; TYPE4SSCAGA.
SQ SEQUENCE 1179 AA; 13166 MW; 7D0ADC34E6446434 CRC64;

Query Match

Best Local Similarity 87.9%; Score 990.5; DB 2; Length 1179;
Matches 206; Conservative 5; Mismatches 17; Indels 33; Gaps 1;

QY 1 KNGKNDPSKYTOAKSDLENSVDVYINOKYTDKVDNLQAVSAKATGDSFSEVQALAD 60
DB 750 KNGKNDPSKYTOAKSDLENSVDVYINOKYTDKVDNLQAVSAKATGDSFSEVQALAD 60
QY 61 LKNFSKEQLAQOAKNESLNAKKESEIYOSVKNVNGTIVNGLSQAEATTLSSKNFSDIK 120
DB 810 LKNFSKEQLAQOAKNESLNAKKESEIYOSVKNVNGTIVNGLSQAEATTLSSKNFSDIK 120
QY 121 KEINAKLGNFNNNNNGKLNKEPIYAKVKKKGAQASLEPIYAKVKKVAKIDRLNOL 180
DB 870 KEINAKLGNFNNNNNGKLNKEPIYAKVKKKGAQASLEPIYAKVKKVAKIDRLNOL 180
QY 181 ASGLGVVGOA-----AGFPLKRDHKKVDLSKY 207
DB 930 ASGLGVVGOA-----AGFPLKRDHKKVDLSKY 207
QY 208 GLSRNOELAKIDINQAVSE 228
DB 990 GLSRNOELAKIDINQAVSE 1010

RESULT 13
Q6VRG7 HELPY PRELIMINARY; PRT: 1182 AA.
AC Q6VRG7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
GN Name=caga; ORFNames=HP0547;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OX NCBI_Taxid=210;

RESULT 14

Q8RRY0 HELPY PRELIMINARY; PRT: 394 AA.
AC Q8RRY0;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
GN Name=caga;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OX NCBI_Taxid=210;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=ATCC51407;
RA Yamaoka Y., Gutierrez O., Saitou N., Kodama T., Kim J.G., Kashima K.,
Ramirez F.C., Mahachai V., Oatso M.S., Graham D.Y.;
RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB057003; BAE81770.1; -; Genomic DNA.
DR InterPro: IPR005169; Caga.
DR Pfam: PF03507; Caga_1.
DR InterPro: IPR004355; IYSec_caga.
DR PRINTS: PRO1553; TYPE4SSCAGA.
SQ SEQUENCE 1182 AA; 131615 MW; BDEC316BE4F1E27 CRC64;

Query Match

Best Local Similarity 87.3%; Score 984; DB 2; Length 1182;
Matches 207; Conservative 5; Mismatches 16; Indels 36; Gaps 2;

QY 1 KNGKNDPSKYTOAKSDLENSVDVYINOKYTDKVDNLQAVSAKATGDSFSEVQALAD 60
DB 750 KNGKNDPSKYTOAKSDLENSVDVYINOKYTDKVDNLQAVSAKATGDSFSEVQALAD 60
QY 61 LKNFSKEQLAQOAKNESLNAKKESEIYOSVKNVNGTIVNGLSQAEATTLSSKNFSDIK 120
DB 810 LKNFSKEQLAQOAKNESLNAKKESEIYOSVKNVNGTIVNGLSQAEATTLSSKNFSDIK 120
QY 121 KEINAKLGNFNNNNNGKLNKEPIYAKVKKKGAQASLEPIYAKVKKVAKIDRLNOL 178
DB 870 KEINAKLGNFNNNNNGKLNKEPIYAKVKKKGAQASLEPIYAKVKKVAKIDRLNOL 178
QY 179 QIASGLGVVGOAAGFPLKRDH-----KVDL 204
DB 930 QIASGLGVVGOAAGFPLKRDH-----KVDL 204
QY 205 SKVGLSRNOELAKIDINQAVSE 228
DB 990 SKVGLSRNOELAKIDINQAVSE 1013

QY 1 KNGKNDPSKYTOAKSDLENSVDVYINOKYTDKVDNLQAVSAKATGDSFSEVQALAD 60
DB 750 KNGKNDPSKYTOAKSDLENSVDVYINOKYTDKVDNLQAVSAKATGDSFSEVQALAD 60
QY 61 LKNFSKEQLAQOAKNESLNAKKESEIYOSVKNVNGTIVNGLSQAEATTLSSKNFSDIK 120
DB 810 LKNFSKEQLAQOAKNESLNAKKESEIYOSVKNVNGTIVNGLSQAEATTLSSKNFSDIK 120
QY 121 KEINAKLGNFNNNNNGKLNKEPIYAKVKKKGAQASLEPIYAKVKKVAKIDRLNOL 178
DB 870 KEINAKLGNFNNNNNGKLNKEPIYAKVKKKGAQASLEPIYAKVKKVAKIDRLNOL 178
QY 179 QIASGLGVVGOAAGFPLKRDH-----KVDL 204
DB 930 QIASGLGVVGOAAGFPLKRDH-----KVDL 204
QY 205 SKVGLSRNOELAKIDINQAVSE 228
DB 990 SKVGLSRNOELAKIDINQAVSE 1013

Query Match

Best Local Similarity 87.2%; Score 982.5; DB 2; Length 394;
Matches 211; Conservative 5; Mismatches 10; Indels 55; Gaps 3;

Search completed: March 6, 2006, 20:29:47
Job time : 231 secs

```
Db 69 KNGKNDPSKVTQAKSDLENSVXDVIINOKITDKVDNLNOAVSAKATGDFSWVEQALAD 128
QY 61 LKNFSKEQLAQQAOKNESLNAKKSEIYOSVKGNGVGTGVNGLSQAETTLTKNPSDIX 120
Db 129 LNNFSKEQLAQQAOKNESLNAKKSEIYOSVKGNGVGTGVNGLSQAETTLTKNPSDIX 188
QY 121 KEINAKLGNFNNNNNGLNK-----EPIYAKVKKKAGQAASLEBP 161
Db 189 KEINAKLGNFNNNNNR--LKNEPIYAKVKKKAGQAASPEEPIYAKVKKKAGQAASPEEP 246
QY 162 IYQAVAKKVAKIDRLNQIASGLGVGQAAGFPLKRHDKYDLS----- 205
Db 247 IYQAVAKKVAKIDRLNQIASGLGVGQAAGFPLKRHDKYDLSKVGLSARSEPIYATID 306
QY 206 -----KVGLSRNOELAKIDNLNOAVSE 228
Db 307 GLGGPPPLKRHDKYDELIRKGLSRTOQLAKIDNLNOAVSE 347
```

```
RESULT 15
Q8KZB0_HELPY PRELIMINARY; PRT; 327 AA.
AC Q8KZB0;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DE 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Caga (Fragment).
GN Name=caga;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxId=210;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OK181;
RX MEDLINE=22295025; PubMed=12391297; DOI=10.1073/pnas.222375399;
RA Higaehi H., Teutsuni R., Fujita A., Yamazaki S., Asaka M., Azuma T.,
RA Hatakeyama M.;
RT "Biological activity of the Helicobacter pylori virulence factor Caga
is determined by variation in the tyrosine phosphorylation sites.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14428-14433(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OK181;
RA Yamakawa A., Yamazaki S., Azuma T.;
RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB090151; BAC10497.1; -; Genomic_DNA.
DR InterPro; IPR005169; Caga.
DR Pfam; PF03507; Caga; 1.
FT NON_TER 1
FT NON_TER 327
SQ SEQUENCE 327 AA; 35798 MW; 1309C5E2418C7DEC CRC64;
```

Query Match 86.3%; Score 972.5; DB 2; Length 327;
Best Local Similarity 86.4%; Pred. No. 7.6e-50;
Matches 197; Conservative 10; Mismatches 20; Indels 1; Gaps 1;

```
QY 1 KNGKNDPSKVTQAKSDLENSVXDVIINOKITDKVDNLNOAVSAKATGDFSRVEQALAD 60
Db 59 KNGKNDPSKVTQAKSDLENSVXDVIINOKITDKVDNLNOAVSIANSTDFSGVEQALAD 118
QY 61 LKNFSKEQLAQQAOKNESLNAKKSEIYOSVKGNGVGTGVNGLSQAETTLTKNPSDIX 120
Db 119 LKNFSKEQLAQQAOKNESLNAKKSEIYOSVKGNGVGTGVNGLSQAETTLTKNPSDIX 178
QY 121 KEINAKLGNFNNNNNGLNK-----EPIYAKVKKKAGQAASLEBP 161
Db 179 KEINAKLGNFNNNNNGLNK-----EPIYAKVKKKAGQAASLEBP 238
QY 181 ASGLGVGQAAGFPLKRHDKYDLSKVGLSRNOELAKIDNLNOAVSE 228
Db 239 ASGLGVGQAAGFPLKRHDKYDLSKVGLSRNOELAKIDNLNOAVSE 285
```

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2006, 20:41:53 ; Search time 21 Seconds

(without alignments)
217.147 Million cell updates/sec

Title: US-09-360-685C-25

Perfect score: 1127
Sequence: 1 KNGKNDPFSKYTQAKSDLEN.....LSRNOELAQKIDNMQAVSE 228

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 135346 seqs, 20000420 residues

Total number of hits satisfying chosen parameters: 135346

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Published Applications AA New:*

1: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
5: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB pep.*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1127	100.0	1147	US-10-615-668-5	Sequence 5, Appli
2	896.5	79.5	1167	US-11-052-554A-121	Sequence 121, App
3	127	11.3	1290	US-10-485-517-141	Sequence 141, App
4	125	11.1	1448	US-10-485-517-212	Sequence 212, App
5	120.5	10.7	5024	US-10-793-626-2964	Sequence 2964, Ap
6	118	10.5	1713	US-10-766-317-2	Sequence 2, Appli
7	118	10.5	1724	US-10-766-317-6	Sequence 6, Appli
8	118	10.5	1724	US-10-766-317-4	Sequence 4, Appli
9	117	10.4	693	US-11-196-475-68	Sequence 68, Appli
10	115.5	10.2	1095	US-10-793-626-3154	Sequence 3154, Ap
11	114	10.1	3712	US-11-019-711-48	Sequence 48, Appli
12	112	9.9	3712	US-11-019-711-51	Sequence 51, Appli
13	111.5	9.9	700	US-11-196-475-66	Sequence 66, Appli
14	110.5	9.8	708	US-11-196-475-76	Sequence 76, Appli
15	110.5	9.8	739	US-11-087-099-12273	Sequence 12273, A
16	109.5	9.7	2087	US-11-075-185-28	Sequence 28, Appli
17	109	9.7	611	US-10-793-626-2586	Sequence 2586, Ap
18	109	9.7	2480	US-10-995-561-825	Sequence 825, App
19	109	9.7	3116	US-10-995-561-826	Sequence 826, App
20	108	9.6	401	US-11-097-749-3	Sequence 3, Appli
21	108	9.6	863	US-11-097-749-2	Sequence 74, Appli
22	105.5	9.4	700	US-11-196-475-74	Sequence 74, Appli
23	104.5	9.3	1404	US-10-878-556A-169	Sequence 169, App
24	104.5	9.3	1652	US-10-995-561-663	Sequence 663, App
25	104.5	9.3	1938	US-10-995-561-661	Sequence 661, App

26	104.5	9.3	1938	US-10-995-561-662	Sequence 662, App
27	104.5	9.3	1954	US-10-995-561-660	Sequence 660, App
28	104.5	9.3	1972	US-10-995-561-664	Sequence 664, App
29	104.5	9.3	1972	US-10-995-561-666	Sequence 666, App
30	103	9.1	440	US-10-976-933-2	Sequence 2, Appli
31	103	9.1	440	US-10-976-933-4	Sequence 4, Appli
32	102.5	9.1	1279	US-10-793-626-3188	Sequence 3188, Ap
33	102	9.1	663	US-11-196-475-70	Sequence 70, Appli
34	102	9.1	663	US-11-196-475-78	Sequence 78, Appli
35	101.5	9.0	380	US-11-144-833-13	Sequence 13, Appli
36	101	9.0	2036	US-11-124-368A-276	Sequence 276, App
37	101	9.0	2036	US-11-124-368A-281	Sequence 281, App
38	101	9.0	2036	US-11-124-368A-281	Sequence 281, App
39	101	9.0	2044	US-11-124-368A-278	Sequence 278, App
40	101	9.0	2107	US-10-995-561-827	Sequence 827, App
41	101	9.0	2144	US-11-124-368A-277	Sequence 277, App
42	100	8.9	794	US-10-793-626-1050	Sequence 1050, Ap
43	98.5	8.7	903	US-10-689-742-142	Sequence 142, App
44	98.5	8.7	1189	US-11-074-176-134	Sequence 134, App
45	98	8.7	239	US-10-485-517-249	Sequence 249, App

ALIGNMENTS

RESULT 1
US-10-615-668-5
Sequence 5, Application US/10615668

Publication No. US20050276819A1

GENERAL INFORMATION:

APPLICANT: Covacci, Antonello

APPLICANT: Bugnoli, Massimo

APPLICANT: Telford, John

APPLICANT: Macchia, Giovanni

APPLICANT: Rappuoli, Rino

TITLE OF INVENTION: Helicobacter Pylori CAI Antigen Polynucleotides

FILE REFERENCE: CHIR0337

CURRENT APPLICATION NUMBER: US/10/615,668

PRIOR FILING DATE: 2003-07-08

PRIOR FILING DATE: 1995-06-06

PRIOR FILING DATE: 1994-10-21

PRIOR APPLICATION NUMBER: 09/410,835

PRIOR FILING DATE: 1999-10-01

NUMBER OF SEQ ID NOS: 24

SOFTWARE: PatentIn version 3.2

SEQ ID NO 5

LENGTH: 1147

TYPE: PRT

ORGANISM: Helicobacter pylori

US-10-615-668-5

Query Match 100.0%; Score 1127; DB 6; Length 1147;
Best Local Similarity 100.0%; Pred. No. 1.4e-69;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	KNGKNDPFSKYTQAKSDLEN	SVQVITINOKVTQKVNINQAVAAKATGPFSSVEQALAD	60
Db	750	KNGKNDPFSKYTQAKSDLEN	SVQVITINOKVTQKVNINQAVAAKATGPFSSVEQALAD	809
QY	61	LKNFSKEQOLAQAQKNSLSNARKKSEIYQSVKNGVNGTIVGNGLSQAEATTLTKNPSFDIK	120	
Db	810	LKNFSKEQOLAQAQKNSLSNARKKSEIYQSVKNGVNGTIVGNGLSQAEATTLTKNPSFDIK	866	
QY	121	KEINAKLGNVNNNNNNKKEPIYAKVNNKKKAQAASLEPIYAQVAKVNAKIDRLNQI	180	
Db	870	KEINAKLGNVNNNNNNKKEPIYAKVNNKKKAQAASLEPIYAQVAKVNAKIDRLNQI	929	
QY	181	ASGIGVVGQAAGPFLKRDHVDLTKVGLSRNOELAQKIDNMQAVSE	228	
Db	930	ASGIGVVGQAAGPFLKRDHVDLTKVGLSRNOELAQKIDNMQAVSE	977	

Tue Mar 7 12:58:24 2006

us-09-360-685c-25.rapbn

Page 2

```
RESULT 2
US-11-052-554A-121
; Sequence 121, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052, 554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589, 227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 121
; LENGTH: 1167
; TYPE: PRF
; ORGANISM: Helicobacter pylori J99
US-11-052-554A-121
```

```
Query Match
Best Local Similarity 79.5%; Score 896.5; DB 7; Length 1167;
Matches 189; Conservative 7; Mismatches 15; Indels 51; Gaps 2;

QY 1 KNGKNDPSKTYOAKSDLENSVKDVIINQKVTDKVDLNOAVSAKATGDFSGVEQALAD 60
DB 754 KNGKNDPSKTYOAKSDLENSIKDVIINQKVTDKVDLNOAVSAKATGDFSGVEQALAD 813
QY 61 LKNFSKEQIAOQAKNESLARKKEBIYOSVKNQVNGTLVNGSGQAEATTLSPKSPDIK 120
DB 814 LKNFSKEQIAOQAKNEDFTGNSALYOSVKNQVNGTLVNGSGKAEATTLSPKSPDIK 873
QY 121 KELNATLGNFNNNNNNKNEPIYAKVKKKAQGAASLEPIYAKVKKAKIDRLNOI 180
DB 874 KELNATLGNFNNNNNNKNEPIYAKVKKKAQGAASLEPIYAKVKKAKIDRLNOI 916
QY 181 ASGLGVNOQAG-----TEPIYQAKVKKAKIDRLNOI 916
DB 917 ASGLGVNOQAGSFLKRDHDKVDLSKVGLSANHEPIYATIDLCGPFILKRDHDKVDLSK 206
QY 207 VGLSRNOELAKTDINNOAVSE 228
DB 977 VGLSRNOELAKTDINNOAVSE 998
```

```
RESULT 3
US-10-485-517-141
; Sequence 141, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485, 517
; PRIOR FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 141
; LENGTH: 1290
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-10-485-517-141
```

```
Query Match
Best Local Similarity 11.3%; Score 127; DB 6; Length 1290;
Matches 69; Conservative 32; Mismatches 93; Indels 78; Gaps 12;

QY 3 GKXKPSKTYOAKSDLENSVKDVIINQKVTDKVDLNOAVSAKATGDFSGVEQALAD 45
DB 181 GNEQAAVNDADKLRQSIDNATTKQONVNTDASQKCDAYANNAVTIAGGIIDQTSR 240
QY 46 -----KATGDFSGVEQALADLKNEF--SKEQIAOQAKNESLARKKEBIYOSVKNQV 95
DB 241 TLDPVYNQAAQGVSTTKNALNGENILEAKQQAQSGLSLDVILNNAQK-----QVTDOI 296
QY 96 NG-----TLVGN-----GLSQAEATTLSPKSPDIKELNATLGNFNNNN 134
DB 297 NGATTVNANQKQNAQNTAMGNLKOALDKDQATKATVNTFTDQ-----AKQAAVNTAV 353
QY 135 NNGKNEPIYAKVKKKAQGAASLEPIYAKVKKAKIDRLNOIASGLGVNOQAGFP 194
DB 354 TNA-----ENIISKANGNATQAE-----VEQAIKQVNAKAKALNGANN----- 392
QY 195 LKRDHDKVDLSKVGLSRNOELAKTDINNOAVSE 226
DB 393 -VQHADEKATLINSNDLNOQAK-DALKQV 422
```

```
RESULT 4
US-10-485-517-212
; Sequence 212, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485, 517
; PRIOR FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 212
; LENGTH: 1448
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-10-485-517-212
```

```
Query Match
Best Local Similarity 11.1%; Score 125; DB 6; Length 1448;
Matches 51; Conservative 45; Mismatches 77; Indels 62; Gaps 8;

QY 4 KNDPSKTYOAKSDLENSVKDVIINQKVTDKVDLNOAVSAKATGDFSGV-EQALAD-- 60
DB 361 QNOAIDNTTATTEERNAKADLVKAKERAYDILN-----AQTNDVYQIKDQAVADIQ 415
QY 61 -----LKNFSKEQIAOQAKNESLARKKEBIYOSVKNQVNGTLVNGSGQAEATTLSP 113
DB 416 GTTADTYIKVADDELATKANEGKALIAQ-----TADATTE 452
QY 114 KNDPSKTYOAKSDLENSVKDVIINQKVTDKVDLNOAVSAKATGDFSGV-EQALAD-- 173
DB 453 K--EQANQGVDAQLTQGNQ-----IENASQIDVDNPK-DNAIQAIDPIQASTDVTNKR 505
QY 174 IDRLNOIASGLGVNOQAGFPILKRDHDKVDLSKVGLSRNOELAKTDINNOAVSE 228
DB 506 AEILTEW-----QKTEITLNNNETTEERKNDIGPVAAVEE 543
```

RESULT 5

US-10-793-626-2964
; Sequence 2964, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUA48005
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2964
; LENGTH: 5024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (5024)
; OTHER INFORMATION: variable amino acid
US-10-793-626-2964

Query Match 10.7%; Score 120.5; DB 6; Length 5024;
Best Local Similarity 23.2%; Pred. No. 4.7;
Matches 62; Conservative 50; Mismatches 84; Indels 71; Gaps 13;

QY 10 KYTQAKSLDENSVDKI-INCKVTDKVNLNOAVSAVATGDFSRVEQALDKKFSKQ 68
DB 74 KLTAKTETAINALNTLADLN--TPQKRAIKATITATRTDVT-AEOSKAKQINSAMHT 129
QY 69 LAQAQAKNES-----INARKK-----SEIYQSVKGVN--GTLVNGLSQ-AEATTL 112
DB 130 LRQNSIDNESVTNENYINAEPEKQAFTEALNNKKEIVNEQATLDANSINQKQALIT 189
QY 113 SKNPSIDKEL-----NA-----KLGNNNNNNNGLKNEPIYAKVKKKAGQASLE 159
DB 190 TKNALDGEOLRAKENADQEIINTLOLTDQORSEKGLVNS--SQRTTEVASQLAK-- 244
QY 160 EPIYQAVAKKNAKIDRLNOIASGLGVGQAAGF-----PLKRHDK 200
DB 245 -----AKELNKVEQLNLTINGKNQINSKFLINBDANOQAVSALASAEVLKNSQ 297
QY 201 VDDLSKVGLSRNOELAKQIDINLNOAVS 227
DB 298 NPELDKV-----TTEQAINNNINSALIN 318

RESULT 6
US-10-766-317-2
; Sequence 2, Application US/10766317
; Publication No. US20060009630A1
; GENERAL INFORMATION:
; APPLICANT: Marinkovich, M. Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING SQUAMOUS CELL CARCINOMA
; FILE REFERENCE: 33828/US/RFT/RMS
; CURRENT APPLICATION NUMBER: US/10/766,317
; CURRENT FILING DATE: 2004-01-27
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In version 3.2
; SEQ ID NO 2
; LENGTH: 1713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-766-317-2

Query Match 10.5%; Score 118; DB 6; Length 1713;
Best Local Similarity 21.2%; Pred. No. 1.8;
Matches 55; Conservative 47; Mismatches 107; Indels 50; Gaps 9;

QY 5 NKDPSKYTQAKSLDENSVDKIINCKVTDKVNLNOAVSAVATGDFSR-----VEQALAD 60
DB 300 NNNVNRATQSAKELDVTKIVNRVHILKQISGTDEGNVPSGDFSRVAAEQRMRE 359
QY 61 L--KNFSKEQLAQAQAKNESIINARKKSEIYQSVKGVNGLVNGLSQAEATTLK--N 115
DB 360 LRNNFNGHLREADREESQLLNRLRTWQKHGEN-----NGLANSIRDSINREYAK 414
QY 116 FSDIKKELNALKGNFNNNNNGLKNEPIYAKVKK-----KAGQASLE 160
DB 415 LSDLRARLOEAAQAQAKNGINQENRALGAIQRYVEINSLQSDFTKYLTTADSSILQT 474
QY 161 PIYQAVAKKNAKIDRLNOIASGLGVGQAAGFLKHKHDVDLS-----KVLGRN----- 212
DB 475 NIALQLEKSOKEYEK-----AASINLEARQ-----ELSDKVELRSAGKSLVEAEK 525
QY 213 -----QELAKQIDINLNOAVS 227
DB 526 ARSLQELAKQLEIKRNAS 544

RESULT 7
US-10-766-317-6
; Sequence 6, Application US/10766317
; Publication No. US20060009630A1
; GENERAL INFORMATION:
; APPLICANT: Marinkovich, M. Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING SQUAMOUS CELL CARCINOMA
; FILE REFERENCE: 33828/US/RFT/RMS
; CURRENT APPLICATION NUMBER: US/10/766,317
; CURRENT FILING DATE: 2004-01-27
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In version 3.2
; SEQ ID NO 6
; LENGTH: 1724
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-766-317-6

Query Match 10.5%; Score 118; DB 6; Length 1724;
Best Local Similarity 21.2%; Pred. No. 1.8;
Matches 55; Conservative 47; Mismatches 107; Indels 50; Gaps 9;

QY 5 NKDPSKYTQAKSLDENSVDKIINCKVTDKVNLNOAVSAVATGDFSR-----VEQALAD 60
DB 311 NNNVNRATQSAKELDVTKIVNRVHILKQISGTDEGNVPSGDFSRVAAEQRMRE 370
QY 61 L--KNFSKEQLAQAQAKNESIINARKKSEIYQSVKGVNGLVNGLSQAEATTLK--N 115
DB 371 LRNNFNGHLREADREESQLLNRLRTWQKHGEN-----NGLANSIRDSINREYAK 425
QY 116 FSDIKKELNALKGNFNNNNNGLKNEPIYAKVKK-----KAGQASLE 160
DB 426 LSDLRARLOEAAQAQAKNGINQENRALGAIQRYVEINSLQSDFTKYLTTADSSILQT 485
QY 161 PIYQAVAKKNAKIDRLNOIASGLGVGQAAGFLKHKHDVDLS-----KVLGRN----- 212
DB 486 NIALQLEKSOKEYEK-----AASINLEARQ-----ELSDKVELRSAGKSLVEAEK 536
QY 213 -----QELAKQIDINLNOAVS 227
DB 537 ARSLQELAKQLEIKRNAS 555

RESULT 8
US-10-766-317-4
; Sequence 4, Application US/10766317
; Publication No. US20060009630A1
; GENERAL INFORMATION:
; APPLICANT: Marinkovich, M. Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING SQUAMOUS CELL CARCINOMA
; FILE REFERENCE: 33828/US/RFT/RMS
; CURRENT APPLICATION NUMBER: US/10/766,317

Tue Mar 7 12:58:24 2006

us-09-360-685c-25.rapbn

Page 4

```

; CURRENT FILING DATE: 2004-01-27
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 4
; LENGTH: 3333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-766-317-4

```

```

Query Match
Best Local Similarity 10.5%; Score 118; DB 6; Length 3333;
Matches 55; Conservative 47; Mismatches 107; Indels 50; Gaps 9;

```

```

QY 5 NKPSKVTOAKSDLENSVQVITINOKVTVDKVDNLNOAVSAKATGDSR-----VEQALAD 60
DB 1920 NNNVNRATOSAKELDVYKIKVIRNVHILKQISGTGEGNNVPSGDSREMAEAGRMRE 1979
QY 61 L--KNFSKEQLAQQAQNSLNAKKSEIYQSVKNGVNGTLVNGLSQAEATTLK---N 115
DB 1980 LRRNNGHLEAALADRRSSQLLRIRTRQKHQGEN-----NGLANSIRDSINEYEAK 2034
QY 116 FSDIKKEINATLGNFNNNNNGLNKNEPIYAKVVK-----KAGQASLEAK 160
DB 2035 LSDLRABLRQMAAQAQNGLNQENBERALGAIORQVEKINSLSGDTKYLTTADSSILQT 2094
QY 161 PIYAQVAKVNAKIDNLNOASGLGVGQAAGFPLKHDVDDLS---KGLSRN--- 212
DB 2095 NIALQWESKQKEYEL---AASINBARQ-----ELSDKVELSRSAKTSIVEAEKH 2145
QY 213 ----OELAKIDNLNOAVS 227
DB 2146 ARSLQELAKQLEIRKNNAS 2164

```

```

RESULT 9
US-11-196-475-68
; Sequence 68, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; PRIOR FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: PacSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 693
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-68

```

```

Query Match
Best Local Similarity 10.4%; Score 117; DB 7; Length 693;
Matches 57; Conservative 53; Mismatches 91; Indels 62; Gaps 12;

```

```

QY 7 DESKVTOAKSDLENS--VVDVYINQKVTVDKVDNLNOAVSAKATGDFS-----RVQO 56
DB 301 DLDRAQQLDSSSDNLDIQRTVYAEKIQIDIEIKENKLPK-PDVSVPVVDKQIQIKE 359

```

```

QY 57 ALADLNKFSKEQLAQQAQNSLNAKKSEIYQSVKN-----GVNGTLVNGLSQ 106
DB 360 LSLDL-----QSLKETSIDENQREIEKQIEKISDEILSKDPKALDNLGDL--NSKVS 413
QY 107 AEATTLISKNFSDIKKEINAKLGNFNNNNNGLNKNEPIYAKVNAKKKAGQAASLEPIYAOV 166
DB 414 SKKKIKGEIYVESKASLADLNDEN---LMEPEDQLSDDKDLKSLKPLFV----- 466
QY 167 AKKVAKIDNLNOIA-SGLGVGQAAGFPLKHDVDDLSKVL----- 209
DB 467 ----SEIEVNEISKSNNEISS--PLYPSISDMSKEDIDNKVDNLQETKSQTKS 519
QY 210 ---SRNQL-AQKIDNLNOAVSE 228
DB 520 QPISLMDLTTMSIDSNSPVYLE 542

```

```

RESULT 10
US-10-793-626-3154
; Sequence 3154, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3154
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-3154

```

```

Query Match
Best Local Similarity 10.2%; Score 115.5; DB 6; Length 1095;
Matches 52; Conservative 33; Mismatches 78; Indels 53; Gaps 9;

```

```

QY 29 QKVTVDVNDLNOAV-----SVAKATGDSRVBQALADLKN-----FSKEQLAQ---QAQ 74
DB 14 OKAEQVIOVANKKIEENAGFSVQVSDSKSVQALSEINNAKSLADADQELQAHVQLT 73
QY 75 KNSLNAKKSEI-----YQSVKNGVNGTLVNGLSQAEATTLISKNFSDIKKEINAKG 128
DB 74 QPTDLNNKPKASITANQRTQOFSNELNST-----KNTDRILKQNSVA 119
QY 129 NFNNNNNNGLNKNEPIYAKVNAKKKAGQAASLEPIYAOVAKVNAKIDNLNOIASGLGVG 188
DB 120 DVNNALN---KRAEVOQKLEAREALLQNKEDNSALVAKXQLQAVD---QVSTGKMQ 173
QY 189 QAAQFPPLKHDVDDLSKVLGRNQLA---OKIDN 221
DB 174 QTK-----DYNSKQAQAQOHSKQAQOVIDN 199

```

```

RESULT 11
US-11-019-711-48
; Sequence 48, Application US/11019711
; Publication No. US20060009634A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Alabrook II, John P
; APPLICANT: Tchernov, Vellizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Patuturajan, Meera

```

APPLICANT: Grose, William M
APPLICANT: Lepley, Denise M
APPLICANT: Burgess, Catherine E
APPLICANT: Vernet, Corine A.M.
APPLICANT: Li, Li
APPLICANT: Gorman, Linda
APPLICANT: Edinger, Shlomit R
APPLICANT: Sciore, Paul
APPLICANT: Ellerman, Karen
APPLICANT: Malenkar, Uriel M
APPLICANT: Rothenberg, Mark
APPLICANT: Stone, David J
APPLICANT: Boldog, Ferenc L
APPLICANT: Guo, Xiaojia
APPLICANT: Shenoy, Suresh G
APPLICANT: Anderson, David W
APPLICANT: Padigaru, Muralidhara
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Miller, Charles E
APPLICANT: Bisen, Andrew J
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-235
CURRENT APPLICATION NUMBER: US/11/019,711
CURRENT FILING DATE: 2004-12-21
PRIOR APPLICATION NUMBER: US/10/037,417
PRIOR FILING DATE: 2002-09-20
PRIOR APPLICATION NUMBER: 60/260,018
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: 60/260,360
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: 60/272,411
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: 60/272,817
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: 60/291,186
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: 60/303,231
PRIOR FILING DATE: 2001-07-05
PRIOR APPLICATION NUMBER: 60/305,060
PRIOR FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: 60/318,405
PRIOR FILING DATE: 2001-09-10
PRIOR APPLICATION NUMBER: 60/318,700
PRIOR FILING DATE: 2001-09-12
NUMBER OF SEQ ID NOS: 227
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 48
LENGTH: 3712
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-11-019-711-48

Query Match 10.1%; Score 114; DB 7; Length 3712;
Best Local Similarity 20.7%; Pred. No. 8.9;
Matches 67; Conservative 46; Mismatches 100; Indels 110; Gaps 15;
5 NNDPSVYTAQKSLNLSVQV---ITN---OKVTDKVNTNOAVS----- 43
2348 NSEKFTVSOQKQAEKNIDNGLNGLTLNQINQKLDNLDALNEINSFNKNVDEEL 2407
44 -----VAKATGDSRVEQALADLNKFSKQQLAQQAQKNBSLN-ARKKSEIYGVKNG 94
2408 PVREDQKKAADLTB--QAEQKAAELAIKQADLAQYTTMTASAEPAITAAATVSGIVEA 2465
95 VNG-----TLVNG-----LSQAEATTLSKNFSDIKKEINA 125
2466 VEAAGKLSQDAISAAGNATDKTDGIBERAHADYTGSTDLLQARQSLQVQDDLEFRLNA 2525
126 -----KLGNNNNNNNGKNEPIYAKVKK-----KKAQAASLE--- 159
2526 SAGKYOKISAVNNATHEHQLKD-----INKLIDQLPAESQDRMWNKSNANASDALEILKN 2579
160 -----EPIYAOVAKK-----VNAKIDRLNG-IASGLGVVGAAGPFLRHKVNDLS--- 205

Db 2580 VAEILEPVSQVTKKELEKAGCIRRDLDLTNKDVQANKQJDDVESSVKLNLEAEDEEQ 2639
Qy 206 --RVGLSRNOELAQKIDNNOAV 226
Db 2640 QHRVG--SQSRQLGQETENLKAQV 2661

RESULT 12
US-11-019-711-51
Sequence 51, Application US/11019711
Publication No. US20060009634A1
GENERAL INFORMATION:
APPLICANT: Kekuda, Ramesh
APPLICANT: Alsobrook II, John P
APPLICANT: Tchernyev, Velizar T
APPLICANT: Liu, Xiaohong
APPLICANT: Spytek, Kimberly A
APPLICANT: Baturajan, Meera
APPLICANT: Grose, William M
APPLICANT: Lepley, Denise M
APPLICANT: Burgess, Catherine E
APPLICANT: Vernet, Corine A.M.
APPLICANT: Li, Li
APPLICANT: Gorman, Linda
APPLICANT: Edinger, Shlomit R
APPLICANT: Sciore, Paul
APPLICANT: Ellerman, Karen
APPLICANT: Malenkar, Uriel M
APPLICANT: Rothenberg, Mark
APPLICANT: Stone, David J
APPLICANT: Boldog, Ferenc L
APPLICANT: Guo, Xiaojia
APPLICANT: Shenoy, Suresh G
APPLICANT: Anderson, David W
APPLICANT: Padigaru, Muralidhara
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Miller, Charles E
APPLICANT: Bisen, Andrew J
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-235
CURRENT APPLICATION NUMBER: US/11/019,711
CURRENT FILING DATE: 2004-12-21
PRIOR APPLICATION NUMBER: US/10/037,417
PRIOR FILING DATE: 2002-09-20
PRIOR APPLICATION NUMBER: 60/260,018
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: 60/260,360
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: 60/272,411
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: 60/272,817
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: 60/291,186
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: 60/303,231
PRIOR FILING DATE: 2001-07-05
PRIOR APPLICATION NUMBER: 60/305,060
PRIOR FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: 60/318,405
PRIOR FILING DATE: 2001-09-10
PRIOR APPLICATION NUMBER: 60/318,700
PRIOR FILING DATE: 2001-09-12
NUMBER OF SEQ ID NOS: 227
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 51
LENGTH: 3712
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-11-019-711-51
Query Match 9.9%; Score 112; DB 7; Length 3712;
Best Local Similarity 21.1%; Pred. No. 12;

Tue Mar 7 12:58:24 2006

us-09-360-685c-25.rapbn

Page 6

Matches	66;	Conservative	41;	Mismatches	104;	Indels	110;	Gaps	14
Qy	5	NNDPSTKQAQKSLDLSNVYEDV---	IIN---	QKVTGRKVDININQVVS---					
Db	2348	NKPTPTYSQKLOAEKNTKIDAGNELINQDITLINOINKDINIRDLNLSFPKAVBEL		2407					
Qy	44	-----VAKATGSPSVGEALADI-KNFSEKQLAQQAQKESINL-ARKSEIYGSVANG		94					
Db	2408	PVREDOHKEDADLTLD--QABQKAEALAKQADLQAQDITMTASAEPIPTAAATAYSGVIEA		2465					
Qy	95	VNG-----TLVGNQ---	LSQAEITLISKNPSDIKELINA	125					
Db	2466	VEAAOKLSODATISAAGNATKRTGDIIEBAAHLADPTGTDILGRANQSLQYVDIEEPYLA		2522					
Qy	126	-----KLGSPNNNNNNNGELKNIEPIYAKYNK-----							
Db	2526	SAGKVKQKISAVNNATTEHQLD-----INKLIDQLPAESQCDHWIKSNNSNNSADLIELIK		2579					
Qy	160	-----BEIYAQVAKK-----VNAKIDELNQAISGLG-----VYGAQGFAPKSHIKVND		203					
Db	2580	VLEIETPEPVAVQPTPELEKAGINRDLIDITKNDVQAANKQDLDDVEGVSLSLSELAIEDIEBO		2639					
Qy	204	LSKVGSLSRNQSLOAKITININQAV		226					
Db	2640	QHRVG--SGSRQLGQELIENKAGV		2661					

```

USUTIT 13
US-11-196-475-66
Sequence 66, Application US/11196475
Publication No. US20050271682A1
GENERAL INFORMATION:
APPLICANT: Dattwyler, Raymond J.
APPLICANT: Gomes Solecki, Maria J. C.
APPLICANT: Luft, Benjamin J.
APPLICANT: Dunn, John J.
TITLE OF INVENTION: Recombinant Constructs of Borrelia
FILE REFERENCE: 2631.1001-01
CURRENT APPLICATION NUMBER: US/11/196,475
PRIOR FILING DATE: 2005-08-03
PRIOR APPLICATION NUMBER: US 08/148,191
PRIOR FILING DATE: 1993-11-01
PRIOR APPLICATION NUMBER: US 08/235,836
PRIOR FILING DATE: 1994-04-29
PRIOR APPLICATION NUMBER: US 09/666,017
PRIOR FILING DATE: 2000-09-19
PRIOR APPLICATION NUMBER: US 60/226,484
PRIOR FILING DATE: 2000-08-18
PRIOR APPLICATION NUMBER: PCT/US01/24736
NUMBER OF SEQ ID NOS: 213
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 66
LENGTH: 700
TYPE: PRT
ORGANISM: Borrelia burgdorferi
US-11-196-475-66

```

```

Query Match      9.9%, Score 11.5; DB 7; length 700;
Best Local Similarity 19.4%; Prid. No. 1.6;
Matches 50; Conservative 59; Mismatches 68; Indels 81; Gaps 10;

QY      5 NKPESKYTAKSDLENSGVKIVINQKTVDKYNLNOAVSVAKAGDPS-----RVGE 56
Db       303 DKAQQKLDAEDNND--VQRNTVRKKIQEDENIINKNLKR- PGVSPPSKVDQLQIRE 359
          :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY      57 ALADLNPFKQAQAQAKRKESILARKSSIKSYYSXNGVNGTLVNGLSQAQATLLSKNF 116
Db       360 SLIEDL---QGLKMGTDGDKRERIEGRQIRIKSEK-----LKSCKDRASKKG 405
          :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY     117 S-DIKELANK-----LGNNNNNNNNGIKNE---PIYAKVK 149
          |   |   |   |   |   |   |   |   |   |   |   |   |   |

```

Db 406 KALDREINSAKSEKSKAKEBEITIGKSGÖKSLGDLDNNENLMPEDOKLPEYKKLDS 465
Qy 150 KKAGQASALBEPYIAQVAKKVNAKTIDRLNOASGLGVGOAAGPLLRKHDKYDLSKVL 209
Db 466 KKEKFEVVS-----EVEKDKLT-----FASNNVNGELSLPLDK 496
Qy 210 SRNOELAQKIDNLNOQAVS 227
Db 497 SSYKQIDSKKEETVKNQVN 514

```

1      RESULT 14
2      US-11-196-475-76
3      / Sequence 76 Application US/11196475
4      / Publication No US20050271682A1
5      / GENERAL INFORMATION:
6      / APPLICANT: Dattagupta, Raymond J.
7      / APPLICANT: Gomes, Stacey, Maria J. C.
8      / APPLICANT: Luft, Benjamin J.
9      / APPLICANT: Dunn, John W.
10     / TITLE OF INVENTION: Recombinant Constructs of Borrelia
11     / TITLE OF INVENTION: Recombinant Constructs of Borrelia
12     / FILE REFERENCE: 2631.1001-011
13     / CURRENT APPLICATION NUMBER: US/11196,475
14     / CURRENT FILING DATE: 2005-08-03
15     / PRIOR APPLICATION NUMBER: US 08/148,191
16     / PRIOR FILING DATE: 1993-11-01
17     / PRIOR APPLICATION NUMBER: US 08/225,836
18     / PRIOR FILING DATE: 1994-04-29
19     / PRIOR APPLICATION NUMBER: US 09/666,017
20     / PRIOR FILING DATE: 2000-09-19
21     / PRIOR APPLICATION NUMBER: US 06/226,484
22     / PRIOR FILING DATE: 2000-08-18
23     / PRIOR APPLICATION NUMBER: PCT/US01/24736
24     / PRIOR FILING DATE: 2001-08-07
25     / NUMBER OF SEQ ID NOS: 213
26     / SOFTWARE: PatsSeq for Windows Version 4.0
27     / SEQ ID NO: 76
28     / LENGTH: 706
29     / TYPE: PRT
30     / ORGANISM: Borrelia burgdorferi
31     / US-11-196-475-76

```

```

Query Match 9.8%; Score 110.5; DB 7; Length 708;
Beet Local Similarity 12.4%; Pred. No. 1.9;
Matches 51; Conservative 54; Mismatches 94; Indels 39; Gaps 9;

QY      5 NKDSKVTAKGSDLENSYKCVITINOKYTDKVDNLQAAVSAAKATGDFS-----RVAG 56
Db      303 DKAGQKLDASBEND--VQNDYTRFKTIQEDINEIKENKNLPK-PGVVSPYKQKQIME 359
QY      57 ALADLKPKSKQGLAQAQKQESLNAARK-----SEIYSVAKGNGTTLVAGLSQAA 109
Db      360 SLELD-----QGLVEMGDNRKREIKQEIIRKDEELSKQKQKSVQYEAALDLREIS 415
QY      110 TTLSKSPFSIKKE-----LNKKGNNNNNNGLNKEPIYAKVKKKQAQAASLEPIY 163
Db      416 KASSKSKSVKEBEETIKGKSASLGDLDNKKMLPBD-----QKLEPDKDKD 467
QY      164 AQAVKAVKAKIDRLNLIQIA-SGLGVYQAAGPFLRRHDKYVDLSKVGLSRNGEIAQKID 220
Db      468 GKKEPFEVSEVKLDKITSNNNNEVKKLSPLDKSYDDID--SKREVDNAIINLKID 523

RESULT 15
US-11-087-099-12273
Sequence 12273, Application US/11087099
Publication No. US20060041961A1
GENERAL INFORMATION:
APPLICANT: Abad, Mark S. et al.
TITLE OF INVENTION: Genes and Uses for Plant Improvement
FILE REFERENCE: 38-21(53450) B EP
CURRENT APPLICATION NUMBER: US/11/087_099

```

! CURRENT FILING DATE: 2005-03-22
! NUMBER OF SEQ ID NOS: 12464
! SEQ ID NO 12273
! LENGTH: 739
! TYPE: PRT
! ORGANISM: Streptococcus pneumoniae
US-11-087-099-12273

Query Match 9.8%; Score 110.5; DB 7; Length 739;
Best Local Similarity 23.1%; Pred. No. 2;
Matches 61; Conservative 43; Mismatches 83; Indels 77; Gaps 10;

QY 1 KNGKNDPFSKYQAK--SDLENSVKDVI-INQKVTDKVDNLNOAVSVAKATGDFSR----- 53
DB 365 ENSKKEBELSKTAEELTAPEQPKDTLKEPKYAAAEKKEVEAAKKAQKQKEDRRNYP 424
QY 54 -----VEQALADLKNPSKEQLAQQAQKNSGLNARKKSEIYQSVKNGVNGTLVNGLS 105
DB 425 TNYKTLLEIAESDVKKAELELVKEANESRNEKIKQAKEVES-----K 473
QY 106 QAEATLTKNPSDIKK-ELNAKIGNFNNNNNGLKNEPIYAKVKKKAGQA-ASLEPTIY 163
DB 474 KAEATLLEKIKTRKKAEEBAK-----KAESEKKAABAKQKVDABEY 517
QY 164 AQYAK--KYNAKIDRLNOIASGLGVGQAAGFPLKRDHVD-----DL 204
DB 518 ALEAKIAELEVEYQRIEK-----ELKEIDSDSEDYKKEGLRAPLQSKLDT 563
QY 205 SKVGLSRNOELAQKIDNLNOAVSR 228
DB 564 KKAQKLSKLELSDKIDELDAEIAK 587

Search completed: March 6, 2006, 20:44:57
Job time : 21 secs

This Page Blank (uspto)